

On Average Risk-sensitive Markov Control Processes

Yun Shen^{*} Klaus Obermayer[†] Wilhelm Stannat[‡]

July 22, 2015

Abstract

We introduce the Lyapunov approach to optimal control problems of average risk-sensitive Markov control processes with general risk maps. Motivated by applications in particular to behavioral economics, we consider possibly non-convex risk maps, modeling behavior with mixed risk preference. We introduce classical objective functions to the risk-sensitive setting and we are in particular interested in optimizing the average risk in the infinite-time horizon for Markov Control Processes on general, possibly non-compact, state spaces allowing also unbounded cost. Existence and uniqueness of an optimal control is obtained with a fixed point theorem applied to the nonlinear map modeling the risk-sensitive expected total cost. The necessary contraction is obtained in a suitable chosen seminorm under a new set of conditions: 1) Lyapunov-type conditions on both risk maps and cost functions that control the growth of iterations, and 2) Doeblin-type conditions, known for Markov chains, generalized to nonlinear mappings. In the particular case of the entropic risk map, the above conditions can be replaced by the existence of a Lyapunov function, a local Doeblin-type condition for the underlying Markov chain, and a growth condition on the cost functions.

Keywords. Markov control processes, Poisson equation, risk-sensitive control, risk measures, stability of nonlinear operators, Doeblin condition, Lyapunov stability.

AMS. 60J05, 93E20, 93C55, 47H07, 91B06

1 Introduction

The average cost (or equivalently, ergodic cost) criterion is a popular infinite-time horizon criterion for optimizing stochastic dynamic systems that are typically modeled

^{*}Fakultät Elektrotechnik und Informatik, Technische Universität Berlin, Marchstr. 23, 10587, Berlin, Germany (yun@ni.tu-berlin.de).

[†]Fakultät Elektrotechnik und Informatik, and Bernstein Center for Computational Neuroscience, Technische Universität Berlin, Marchstr. 23, 10587, Berlin, Germany (oby@ni.tu-berlin.de). The work of this author and the first author was supported by the BMBF, Bernstein Fokus Lernen TP1, 01GQ0911.

[‡]Institut für Mathematik, and Bernstein Center for Computational Neuroscience, Technische Universität Berlin, Straße des 17. Juni 136, 10623, Berlin, Germany (stannat@math.tu-berlin.de). The work of this author is supported by the BMBF, FKZ01GQ1001B.

by *Markov control processes* (MCPs, see, e.g., [25, 26], and [39] under the name *Markov decision processes*). The research (see e.g., [1] and [26, Chapter 10] for a comprehensive survey) focuses on the expected long-run average cost, which is a risk-neutral criterion.

The aim of this paper is to consider a general risk-sensitive criterion. So far, most of the long-run average risk-sensitive (especially, risk-averse) control employs the exponential utility function [3, 5, 6, 9, 10, 14, 18, 23, 27, 29, 35].

In recent years, pioneered by Ruszczyński [42], many authors ([4, 7, 45]) have developed a more general framework of risk-sensitive sequential control problems on Borel spaces by applying *coherent/convex risk measures*, which were originally employed in mathematical finance by Artzner et al. [2], Föllmer and Schied [19], to the classical risk-neutral MCPs. In particular, the exponential utility function can be viewed as a special convex (but not coherent) risk measure, called *entropic measure* [19]. Among them, Ruszczyński [42] considered coherent risk measures with the finite-horizon and discounted criteria, Bäuerle and Rieder [4] considered convex risk measures with bounded costs, and Çavuş and Ruszczyński [7] considered coherent risk measure with total undiscounted criterion for transient MCPs. In this paper, we shall apply a more general family of risk measures to ergodic MCPs on Borel state-action spaces equipped with possibly unbounded costs, and solve the optimization problem corresponding to the average risk-sensitive criterion.

In the same setting, Shen et al. [45] introduced the concept of *risk map* that generalizes the one-step conditional risk measure (see e.g., [42]). Weighted norm spaces were then applied to incorporate possibly unbounded costs and Lyapunov-type stability conditions that generalized known conditions for Markov chains were stated to ensure the existence of solutions to the optimality equation for the average risk-sensitive criterion. More specifically, to deal with unbounded costs and nonlinearity of risk maps, [45] introduced a dominating coherent risk map (called *upper module*) and assumed the existence of a Lyapunov function to it. However, given a non-coherent risk map, e.g., the entropic measure, this dominating risk map may be infinite for unbounded cost functions. Hence, a (real-valued) Lyapunov function need not exist and the theory developed in [45] is reduced to be valid only for bounded costs.

In order to solve the above mentioned problem within the same framework, we introduce in this paper 1) constraints both on the cost function and risk map to control the growth of *value iterations* and 2) additional minorization properties on bounded level-sets (small sets) that generalize the standard Doeblin condition. Under these conditions, we show the existence of a *bounded forward invariant subset* that covers the whole iterations. Restricted to the bounded subset, we assume the existence of the Lyapunov function for the weaker type of dominating map, which is called *upper envelope* in this paper, to ensure the existence of a unique solution to the optimality equation for the average-risk criterion. As a special case, we show that, when applying the entropic map, the above conditions are satisfied, if 1) a Lyapunov function exists for the entropic measure, 2) the local Doeblin condition holds for the underlying Markov chain, and 3) a growth condition for cost functions.

Most of the existing literature on risk-sensitive MCPs, especially that applies the entropic map, considers finite or countable state spaces (see, e.g., [3, 5, 6, 10, 18,

23, 35]), or bounded cost functions (see, e.g., [4]). Comparing with the literature [13, 14, 29] of the same settings, i.e., Borel spaces and unbounded cost functions, we provide in this paper a more general framework which can be applied to all types of risk maps, and more importantly, with a conceptually simpler proof, whereas the methods developed in [13, 14, 29] can only be applied to the entropic map. Moreover, the conditions we state for the entropic map are simpler than the conditions stated in [13, 14, 29].

The paper is organized as follows. In Section 2, we briefly review the framework of MCPs and Lyapunov approach for ergodic MCPs with the weighted norm space, followed by an introduction of risk maps within the context of MCPs. In Section 3, the average risk criterion is formally defined and two sets of general conditions are stated to ensure the boundedness of iterations and the geometric contraction. We then prove the existence and uniqueness of a solution to the associated nonlinear Poisson equation, as well as the average risk criterion. As a special case, we show in Section 4 that under proper verifiable assumptions, the entropic map fits the theoretical framework developed in the previous section. Finally, in order to demonstrate the applicability of our theory, we show in Section 5 that applied to a canonical MCP, various types of risk maps fit also the developed theoretical framework.

Notation

Let X be a *Borel space*, that is a Borel subset of a complete separable metric space, and $\mathcal{B}(X)$ its Borel σ -algebra. Let X and Y be two Borel spaces. A *stochastic kernel on X given Y* is a function $\psi(B|y), B \in \mathcal{B}(X), y \in Y$ such that i) $\psi(\cdot|y)$ is a probability measure on $\mathcal{B}(X)$ for every fixed $y \in Y$, and ii) $\psi(B|\cdot)$ is a measurable function on Y for every fixed $B \in \mathcal{B}(X)$. Let $\bar{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$ be the extended real line.

2 Model and problem formulation

2.1 Markov control processes

In this subsection, we briefly introduce the framework of Markov control processes, where we mostly follow the notation by Hernández-Lerma and Lasserre [26]. A Markov control process, $(X, A, \{A(x)|x \in X\}, Q, c)$, consists of the following components: *state space* X and *action space* A , which are Borel spaces; the feasible action set $A(x)$, which is a nonempty Borel space of A , for a given state $x \in X$; the *transition model* $Q(B|x, a), B \in \mathcal{B}(X), (x, a) \in K$: a *stochastic kernel* on X given K , where K denotes the set of feasible state-action pairs $K := \{(x, a)|x \in X, a \in A(x)\}$, which is a Borel subset of $X \times A$; and the *cost function* $c: K \rightarrow \bar{\mathbb{R}}, \mathcal{B}(K)$ -measurable. Random variables are denoted by capital letters, whereas realizations of the random variables are denoted by lowercase letters. The process is assumed to be Markov, i.e., for each $t \in \mathbb{N}$, $\mathbb{P}(X_{t+1} \in B|X_t = x, A_t = a, X_{t-1}, \dots, X_0, A_0) = Q(B|x, a), \forall B \in \mathcal{B}(X)$ and $(x, a) \in K$.

We restrict in this paper to *Markov policies*, $\pi = [\pi_0, \pi_1, \pi_2, \dots]$, where each *single-step policy* $\pi_t(\cdot|x_t)$, which denotes the probability of choosing action a_t at x_t , $(x_t, a_t) \in K$, is Markov (independent of the states and actions before t) and, therefore,

a stochastic kernel on \mathbf{A} given \mathbf{X} . We use the boldface to represent a sequence of policies while using the normal typeface for a single-step policy. Let Δ denote the set of all stochastic kernels on \mathbf{A} given \mathbf{X} , μ , such that $\mu(\mathbf{A}(x)|x) = 1$ and $\Pi_M := \Delta^\infty$ denotes the set of all Markov policies. A policy $f \in \Delta$ is *deterministic* if for each $x \in \mathbf{X}$, there exists some $a \in \mathbf{A}(x)$ such that $f(\{a\}|x) = 1$. Let $\Delta_D \subset \Delta$ denote the set of all deterministic single-step policies. A policy π is said to be *stationary*, if $\pi = \pi^\infty$ for some $\pi \in \Delta$. For each $x \in \mathbf{X}$ and single-step policy $\pi \in \Delta$, define

$$(1) \quad c^\pi(x) := \int_{\mathbf{A}(x)} c(x, a) \pi(da|x), P^\pi(\mathbf{B}|x) := \int_{\mathbf{A}(x)} Q(\mathbf{B}|x, a) \pi(da|x), \mathbf{B} \in \mathcal{B}(\mathbf{X}).$$

The following *average cost* (see, e.g., Arapostathis et al. [1] and Hernández-Lerma and Lasserre [26, Chapter 10]) is used as an objective:

$$(2) \quad S := \limsup_{T \rightarrow \infty} \frac{1}{T} S_T, \text{ where } S_T := \sum_{t=0}^T c(X_t, A_t).$$

The optimization problem is then to minimize the expected objective

$$(3) \quad \inf_{\pi \in \Pi_M} \mathbb{E}^\pi [S | X_0 = x]$$

by selecting a policy π . We notice that due to the Markov property, the finite-stage objective function can be decomposed as follows,

$$(4) \quad \mathbb{E}_{X_0}^\pi [S_T] = c^{\pi_0}(X_0) + \mathbb{E}_{X_0}^{\pi_0} [c^{\pi_1}(X_1) + \mathbb{E}_{X_1}^{\pi_1} [c^{\pi_2}(X_2) + \dots + \mathbb{E}_{X_{T-1}}^{\pi_{T-1}} [c^{\pi_T}(X_T)] \dots]],$$

where $\mathbb{E}_{X_t}^{\pi_t} [v(X_{t+1})]$ denotes the *conditional expectation* of the function v of the successive state X_{t+1} given current state X_t .

2.2 Lyapunov approach for ergodic MCPs

It is known that the optimization problem of the average cost criterion is closely related to the w -ergodicity of underlying Markov processes (see e.g., [26, Chapter 10]). This is usually established by finding a Lyapunov function w with “small” level-sets [36, Chapter 14]. If the Lyapunov function is strong enough, the transition kernels converge exponentially fast towards the unique invariant measure. Among other variations [11, 15, 31], Hairer and Mattingly [22] stated a simplified version of the conditions, which is the main approach that we will follow and extend in this paper.

Weighted norm

We first introduce the weighted norm and seminorm. Let $w : \mathbf{X} \rightarrow [1, \infty)$ be a given real-valued $\mathcal{B}(\mathbf{X})$ -measurable function. Consider the w -norm

$$\|u\|_w := \sup_{x \in \mathbf{X}} \frac{|u(x)|}{w(x)}.$$

Let \mathcal{B}_w be the space of real-valued $\mathcal{B}(\mathbf{X})$ -measurable functions with bounded w -norm. It is obvious that $\mathcal{B} \subset \mathcal{B}_w$, where \mathcal{B} denotes the space of bounded $\mathcal{B}(\mathbf{X})$ -measurable functions. Throughout this paper we call w interchangeably a *weight function* or a *Lyapunov function*. One notable property is as follows: if $w_0 : \mathbf{X} \rightarrow [0, \infty)$ is a $\mathcal{B}(\mathbf{X})$ -measurable function, then $1 + \beta w_0$ is a valid weight function for any positive coefficient β . Furthermore, $\mathcal{B}_{1+\beta_1 w_0} = \mathcal{B}_{1+\beta_2 w_0}$ for any two positive coefficients, β_1 and β_2 . This property will be used several times throughout this paper.

For a given signed measure μ on $\mathcal{B}(\mathbf{X})$ with $\int w d|\mu| < \infty$, the integral $\int u d\mu$ is well-defined for all $u \in \mathcal{B}_w$. Let

$$\|\mu\|_w := \sup_{\|u\|_w \leq 1} \left| \int_{\mathbf{X}} u d\mu \right|$$

and note that

$$\|\mu\|_w = \int_{\mathbf{X}} w d|\mu| \geq \|\mu\|_{TV},$$

where $\|\cdot\|_{TV}$ denotes the total variation norm. Let \mathcal{P} be the space of all probability measures μ on $\mathcal{B}(\mathbf{X})$ and $\mathcal{P}_w := \{\mu \in \mathcal{P} \mid \int w d|\mu| < \infty\}$. For any $\mu \in \mathcal{P}_w$ and $f \in \mathcal{B}_w$, we introduce the following notation

$$\mu[f] := \int_{\mathbf{X}} f(x) \mu(dx).$$

The following w -seminorm is used throughout this paper:

$$\|v\|_{s,w} := \sup_{x \neq y} \frac{|v(x) - v(y)|}{w(x) + w(y)}.$$

When restricting to the space \mathcal{B} , i.e., setting $w \equiv 1$, the seminorm is called *span-norm* in [24] or *Hilbert seminorm* in [21]. The following lemma (see [22, Lemma 2.1]) establishes the connection between the weighted norm and seminorm.

LEMMA 2.1. $\|v\|_{s,w} = \min_{c \in \mathbb{R}} \|v + c\|_w, \forall v \in \mathcal{B}_w$.

The partial ordering “ \leq ” between elements in \mathcal{B}_w is defined as follows: we say $v \leq u$ if $v(x) \leq u(x) \forall x \in \mathbf{X}$. A real number $u \in \mathbb{R}$ can be viewed as a constant-valued function which belongs also to \mathcal{B}_w .

Ergodicity conditions

We adapt the sufficient conditions for ergodicity of Markov chains stated by Hairer and Mattingly [22] to the MCP-framework.

Assumption 2.2. (i) There exists a function $w : \mathbf{X} \rightarrow [0, \infty)$, which is $\mathcal{B}(\mathbf{X})$ -measurable, and constants $K \geq 0$ and $\gamma \in (0, 1)$ such that

$$\int Q(dy|x, a) w(y) \leq \gamma w(x) + K, \forall (x, a) \in \mathbf{K}$$

(ii) There exists a constant $\alpha \in (0, 1)$ and a probability measure μ so that

$$Q(C|x, a) \geq \alpha\mu(C), \forall C \in \mathcal{B}(X), x \in B, a \in A(x)$$

with $B := \{x \in X : w(x) \leq R\}$ for some $R > \frac{2K}{1-\gamma}$.

Here, (i) is a Lyapunov-type condition that controls the growth of iterations with a Lyapunov function w , while (ii) is a Doeblin-type condition (see e.g., [34]) that assumes a support μ uniformly on the bounded level-set B .

Under this assumption, a direct extension of Theorem 3.1 in [22] shows that there exist constants $\bar{\alpha} \in (0, 1)$ and $\beta > 0$, both depending on γ, K and α , such that

$$(5) \quad \|P^\pi[v]\|_{s, 1+\beta w} \leq \bar{\alpha}\|v\|_{s, 1+\beta w}, \forall v \in \mathcal{B}_{1+\beta w}, \pi \in \Delta.$$

This ensures the ergodicity of the underlying MCPs (cf. [26, Chapter 10]) and also guarantees the geometric convergence of value iterations. In Section 3, we shall generalize this set of conditions to nonlinear operators.

2.3 Risk maps

Definition

Inspired by risk measures applied in mathematical finance [2, 19], we introduce in the following our version of risk measures on the weighted norm space \mathcal{B}_w and the probability measure space \mathcal{P} .

DEFINITION 2.3. *A mapping $\nu : \mathcal{B}_w \times \mathcal{P} \rightarrow \bar{\mathbb{R}}$ is said to be a risk measure if it satisfies that for each $\mu \in \mathcal{P}$,*

(i) *(monotonicity) $\nu(v, \mu) \leq \nu(u, \mu)$, whenever $v \leq u \in \mathcal{B}_w$;*

(ii) *(translation invariance) $\nu(v + u, \mu) = \nu(v, \mu) + u$, $\forall u \in \mathbb{R}, v \in \mathcal{B}_w$;*

(iii) *(centralization) $\nu(0, \mu) = 0$.*

A mapping $\nu : \mathcal{B}_w \rightarrow \bar{\mathbb{R}}$ is said to be a risk measure with respect to (w.r.t.) $\mu \in \mathcal{P}$ if there exists a risk measure $\tilde{\nu}$ such that $\nu(v) = \tilde{\nu}(v, \mu)$, $\forall v \in \mathcal{B}_w$. Furthermore, ν is said to be real-valued if $|\nu(v)| < \infty$, $\forall v \in \mathcal{B}_w$.

Remark 2.4. Comparing with the definition of risk measures in [2, 19], we make the following extensions:

1. We consider here the weighted space \mathcal{B}_w , adapted to a given Lyapunov function w , rather than the space of bounded random variables, $L^\infty(\mu)$, since \mathcal{B}_w is more suitable for investigating the stability properties of the underlying Markov process (see, e.g., [22, 36, 45]) and is also more general than $L^\infty(\mu)$. We will show in later sections how to specify w , depending on the form of risk measures and the properties of the underlying Markov process as well.

2. The objective probability measure μ is allowed to vary, since in the framework of MCPs, $Q_{x,a}(\cdot) := Q(\cdot|x, a)$ is a probability measure depending on different state-action pairs, though the transition kernel Q itself is assumed to be fixed and known a priori.
3. We dropped the assumption of coherency or convexity (see Definition 2.6 below for a formal definition), which is necessary for inducing risk-averse behavior. However, studies in behavioral economics (see e.g., prospect theory by Tversky and Kahneman [49]) show that human agents are not always risk-averse. Hence, we drop this assumption and allow more general types of risk measures.

To apply risk measures to MCPs, we follow the approach pioneered by Ruszczyński [42] with, however, more general types of risk measures. We first introduce the concept of risk maps. A similar concept has also been introduced by Çavuş and Ruszczyński [7] for coherency maps, and by Shen et al. [45] for general risk maps.

DEFINITION 2.5. *Let $\{X, A, \{A(x)|x \in X\}, Q, c\}$ be an MCP and $K = \{(x, a)|x \in X, a \in A(x)\}$ be the set of all feasible state-action pairs. A mapping $\mathcal{R}(v|x, a) : \mathcal{B}_w \times K \rightarrow \mathbb{R}$ is said to be a risk map on the MCP if*

- (i) *for each $(x, a) \in K$, $\mathcal{R}(\cdot|x, a) : \mathcal{B}_w \rightarrow \mathbb{R}$ is a real-valued risk measure w.r.t. $Q_{x,a}$, i.e., there exists a real-valued risk measure ν such that $\mathcal{R}_{x,a}(\cdot) = \nu(\cdot, Q_{x,a})$;*
- (ii) *for each $v \in \mathcal{B}_w$, $\mathcal{R}(v|\cdot)$ is a real-valued $\mathcal{B}(K)$ -measurable function.*

Furthermore, we define for any $\pi \in \Delta$, $\mathcal{R}^\pi(v|x) := \int_{A(x)} \pi(da|x) \mathcal{R}(v|x, a)$.

For convenience, we write interchangeably $\mathcal{R}_{x,a}(v) := \mathcal{R}(v|x, a)$ and $\mathcal{R}_x^\pi(v) := \mathcal{R}^\pi(v|x)$. $\mathcal{R}^\pi(v)$ can therefore be viewed as a $\mathcal{B}(X)$ -measurable function for each $v \in \mathcal{B}_w$. It is worth to mention that \mathcal{R} in fact depends on the transition kernel Q of the underlying MCP, though for brevity we omit it in \mathcal{R} , because in the framework of MCPs, Q is assumed to be fixed and known a priori.

Risk preference

We follow the rule that *diversification* should be preferred if the agent is risk-averse (see e.g., [20, Chapter 4]). More specifically, suppose that an agent faces two choices under the same probability measure μ , one of which has a future cost function v while the other v' . If the agent diversifies, i.e., spends a fraction α of the resources on the first and the remaining amount on the other choice, the future cost is given by $\alpha v + (1 - \alpha)v'$. If the applied risk measure is convex, $\rho(\alpha v + (1 - \alpha)v', \mu) \leq \alpha \rho(v, \mu) + (1 - \alpha)\rho(v', \mu)$, $\forall \alpha \in [0, 1]$, then the diversification should reduce the risk. Thus, the agent's behavior is expected to be *risk-averse*. Conversely, if the applied valuation function is *concave*, the induced risk-preference should be *risk-seeking*. This categorization can be extended to risk maps as follows.

DEFINITION 2.6. *A risk map \mathcal{R} on an MCP is said to be*

- *convex, if*

$$\mathcal{R}_{x,a}(\alpha v + (1 - \alpha)v') \leq \alpha \mathcal{R}_{x,a}(v) + (1 - \alpha)\mathcal{R}_{x,a}(v'), \forall \alpha \in [0, 1], (x, a) \in K;$$

- homogeneous, if $\mathcal{R}_{x,a}(\lambda v) = \lambda \mathcal{R}_{x,a}(v), \forall \lambda \in \mathbb{R}_+, (x, a) \in \mathbf{K}$;
- coherent, if it is convex and homogeneous.

Among others, it is easy to verify that coherent risk maps are *subadditive*:

$$(6) \quad \mathcal{R}_{x,a}(v + v') \leq \mathcal{R}_{x,a}(v) + \mathcal{R}_{x,a}(v'), \forall (x, a) \in \mathbf{K}, v, v' \in \mathcal{B}_w.$$

Examples

We list below several important examples that have been widely used in the literature. First of all, it is easy to verify that the conditional expectation

$$(7) \quad \mathcal{R}_{x,a}(v) = Q_{x,a}[v] := \int_{\mathbf{X}} Q(dy|x, a)v(y)$$

is a coherent risk map that is linear to v . In this case, we call the MCP *risk-neutral*.

Example 2.7. Föllmer and Schied [19] introduced the *entropic map*:

$$(8) \quad \mathcal{R}_{x,a}(v) := \frac{1}{\lambda} \ln (Q_{x,a}[e^{\lambda v}]) = \frac{1}{\lambda} \ln \left\{ \int_{\mathbf{X}} Q(dy|x, a) e^{\lambda v(y)} \right\}$$

where the risk-sensitive parameter $\lambda \in \mathbb{R}$ controls the risk-preference of \mathcal{R} : if $\lambda > 0$, \mathcal{R} is everywhere convex and therefore everywhere risk-averse; if $\lambda < 0$, \mathcal{R} is everywhere concave and therefore everywhere risk-seeking. It has been studied intensely also in the field of optimal control [3, 5, 6, 9, 10, 14, 18, 23, 29, 35].

Example 2.8. Iyengar [28] introduced the framework of *robust dynamic programming* (see also [37]), by which he argues that in some applications the transition model Q cannot be inferred exactly. Instead, he employs a set of transition probabilities, \mathcal{P} , which contains all possible “ambiguous” transition kernels. In order to gain the “robustness”, the worst cost is considered, adapted to our framework,

$$(9) \quad \mathcal{R}_{x,a}(v) := \sup_{P(\cdot|x,a) \in \mathcal{P}_{x,a}} P_{x,a}[v] = \sup_{P(\cdot|x,a) \in \mathcal{P}_{x,a}} \int_{\mathbf{Y}} P(dy|x, a)v(y).$$

One example of such $\mathcal{P}_{x,a}$ is given by Ruszczyński and Shapiro [43, Example 4.3]:

$$(10) \quad \mathcal{P}_{x,a} := \left\{ \mu \in \mathcal{P} \mid \mu \ll Q_{x,a}, 0 \leq \gamma_1 \leq \frac{d\mu}{dQ_{x,a}} \leq \gamma_2 < \infty \right\}.$$

Here, “ \ll ” denotes the absolute continuity and Q denotes the true transition kernel of the underlying MCP, while γ_1 and γ_2 control the degree of variation between the estimated transition kernel and its true model. In particular, if $\gamma_1 = 0$, it becomes the *average value at risk* (see [40, 44] and references therein). It is also notable that each concave and homogeneous valuation function has one dual representation of the form (9) under some regularity conditions for the set \mathcal{P} (see e.g. [12] for essentially bounded spaces and [48] for unbounded ones).

Example 2.9. Ogryczak and Ruszczyński [38] considered the trade-off between the one-step conditional mean and semideviation rather than the deviation of the whole Markov chain [17, 47]:

$$(11) \quad \mathcal{R}_{x,a}(v) := Q_{x,a}[v] + \lambda \left(Q_{x,a} \left[(v - Q_{x,a}[v])_+^r \right] \right)^{1/r}$$

where $r \geq 1$ and $\lambda \in [-1, 1]$ denotes the risk-preference parameter which controls the risk preference of \mathcal{R} : if $\lambda > 0$, \mathcal{R} is risk-averse; if $\lambda < 0$, \mathcal{R} is risk-seeking. Here, $(x)_+ := \max(x, 0)$. Setting $r = 2$, this map can be viewed as an approximation of the mean-variance tradeoff scheme defined in [17].

Example 2.10. The entropic map introduced above belongs in fact to a large family of risk maps called *utility-based shortfall* (see [20, Section 4.6] and [44, Section 2]) defined as follows. Let $u : \mathbb{R} \rightarrow \mathbb{R}$ be a continuous, increasing and non-constant utility function satisfying $u(0) = 0$. Assume that there exists a constant $m \in \mathbb{R}$ such that $\int_{\mathbf{X}} u(v(y) - m) Q_{x,a}(\mathrm{d}y) < \infty$ for each $(x, a) \in \mathbf{K}$. Then, the utility-based shortfall is defined as

$$(12) \quad \mathcal{R}_{x,a}(v) := \sup \left\{ m \in \mathbb{R} \mid \int_{\mathbf{X}} u(v(y) - m) Q_{x,a}(\mathrm{d}y) \geq 0 \right\}.$$

It is remarkable that in principle, one can apply different u for different state-action pair (s, a) . Here, for brevity, we apply the same u for all $(s, a) \in \mathbf{K}$.

3 Average risk-sensitive MCPs

We define $\mathcal{T} : \mathcal{B}_w \times \mathbf{K} \rightarrow \mathbb{R}$ as

$$\mathcal{T}(v|x, a) := c(x, a) + \mathcal{R}(v|x, a), (x, a) \in \mathbf{K},$$

and for each single-step policy $\pi \in \Delta$ define

$$(13) \quad \mathcal{T}^\pi(v|x) := \int \pi(\mathrm{d}a|x) \mathcal{T}(v|x, a) = c^\pi(x) + \mathcal{R}^\pi(v|x)$$

which is $\mathcal{B}(\mathbf{X})$ -measurable for each $v \in \mathcal{B}_w$. For convenience, we interchangeably write $\mathcal{T}_{x,a}(v) := \mathcal{T}(v|x, a)$ and $\mathcal{T}_x^\pi(v) := \mathcal{T}^\pi(v|x)$. It is worth to mention that for any $\pi \in \Delta$, monotonicity and translation invariance (but not centralization) (see Definition 2.3) of \mathcal{R} imply the same properties for $\mathcal{T}_x^\pi(\cdot)$.

3.1 Average criterion

Before we construct the risk-sensitive objective using the backward induction as in [42] and [45], we first state conditions, under which $\mathcal{T}^\pi(v|\cdot) \in \mathcal{B}_w$ for fixed $v \in \mathcal{B}_w$ and any $\pi \in \Delta$. For risk-neutral MCPs, where $\mathcal{R}(\cdot) = Q[\cdot]$, this is usually guaranteed by assuming $|c(x, a)| \leq Cw(x)$ and $Q_{x,a}[w] \leq Cw(x)$, for all $(x, a) \in \mathbf{K}$ with some positive constant C (see e.g., [26, Assumption 10.2.1(d) and (f)]). For general risk maps which are not homogeneous, the above assumption is not sufficient. Instead, we consider the following assumption throughout this paper:

(A1) There exist a $\mathcal{B}(\mathbf{X})$ -measurable function $w_0 : \mathbf{X} \rightarrow [0, \infty)$, constants $\gamma_0 \in (0, 1)$ and $K_0 > 0$ such that

$$\mathcal{T}_{x,a}(w_0) \vee (-\mathcal{T}_{x,a}(-w_0)) \leq \gamma_0 w_0(x) + K_0, \forall (x, a) \in \mathbf{K}.$$

Here, $a \vee b := \max(a, b)$. From now on, let $w := 1 + K^{-1}w_0$ with some positive $K \in \mathbb{R}_+$ whose value will be specified in Section 3.2.

PROPOSITION 3.1. *Assume that (A1) holds and $v \in \mathcal{B}_w$ satisfies $|v(x)| \leq w_0(x) + A, \forall x \in \mathbf{X}$, with some positive A . Then $|\mathcal{T}_x^\pi(v)| \leq w_0(x) + A + K_0, \forall x \in \mathbf{X}, \pi \in \Delta$.*

Proof. By assumption, $-w_0 - A \leq v \leq w_0 + A$. (A1) implies that

$$\mathcal{T}_{x,a}(v) \leq \mathcal{T}_{x,a}(w_0) + A \leq \gamma_0 w_0(x) + K_0 + A \leq w_0(x) + K_0 + A.$$

Similarly, we obtain $\mathcal{T}_{x,a}(v) \geq -w_0(x) - K_0 - A$, by which the assertion follows. \square

Let $\boldsymbol{\pi} = [\pi_0, \pi_1, \dots]$ be a Markov policy. The above proposition shows that starting from any $v \in \mathcal{B}_w$ satisfying $|v| \leq w_0 + A$, the following backward iteration

$$v^{\boldsymbol{\pi}, T, T} := \mathcal{T}^{\pi_T}(v), v^{\boldsymbol{\pi}, t, T} := \mathcal{T}^{\pi_t}(v^{\boldsymbol{\pi}, t+1, T}), t = T-1, T-2, \dots, 0$$

is well-defined, since $v^{\boldsymbol{\pi}, t, T} \in \mathcal{B}_w$ for each $t = 0, 1, \dots, T$.

We now apply risk maps to construct risk-sensitive objectives as in [42, 45]. Replacing the conditional expectation in (4) with a risk map \mathcal{R} , we obtain for $\boldsymbol{\pi} \in \Pi_M$ and $x \in \mathbf{X}$,

$$\begin{aligned} (14) \quad J_T(x, \boldsymbol{\pi}) &:= c^{\pi_0}(x) + \mathcal{R}_x^{\pi_0}(c^{\pi_1} + \mathcal{R}^{\pi_1}(c^{\pi_2} + \dots + \mathcal{R}^{\pi_{T-1}}(c^{\pi_T}) \dots)) \\ &= \mathcal{T}_x^{\pi_0}(\mathcal{T}^{\pi_1}(\dots \mathcal{T}^{\pi_T}(0) \dots)), \end{aligned}$$

and the risk-sensitive objective considered in this paper is the *average risk*:

$$(15) \quad J(x, \boldsymbol{\pi}) := \limsup_{T \rightarrow \infty} \frac{1}{T} J_T(x, \boldsymbol{\pi}).$$

Remark 3.2. Applying the same recursive approach as above, other two widely used objectives in literature, the finite-stage total cost and the discounted cost, can be analogously extended to risk-sensitive objectives, the *finite-stage total risk* and the *discounted risk*, respectively (see [42] for coherent and [45] for general risk maps).

Remark 3.3. It is remarkable that the entropic map introduced in Exampe 2.7 is *time-consistent* [32]. To see this in the framework of MCPs, note that for the finite-stage sum,

$$\nu(S_T|x) = \frac{1}{\lambda} \ln \left(\mathbb{E}^\pi [e^{\lambda \sum_{t=0}^T c(X_t, A_t)} | X_0 = x] \right) = \frac{1}{\lambda} \ln \left(\mathbb{E}^\pi \left[\prod_{t=0}^T e^{\lambda c(X_t, A_t)} | X_0 = x \right] \right).$$

Suppose now that the policy is deterministic $\boldsymbol{\pi} = [f_0, f_1, \dots, f_T]$. By the Markov property, we have then

$$\begin{aligned}\nu(S_T|x) &= \frac{1}{\lambda} \ln \left(e^{\lambda c^{f_0}(x)} P_x^{f_0} \left[\dots \left[e^{\lambda c^{f_T}} P^{f_{T-2}} \left[e^{\lambda c^{f_T}} P^{f_{T-1}} [e^{\lambda c^{f_T}}] \right] \right] \dots \right] \right) \\ &= c^{f_0}(x) + \mathcal{R}_x^{f_0} \left(\dots \left(c^{f_{T-1}} + \mathcal{R}^{f_{T-1}}(c^{f_T}) \right) \dots \right).\end{aligned}$$

However, this equivalence between the risk measure and its time-consistent version applied in this paper is an exception rather than the rule (for details see e.g., [41]). It does not hold for the other risk maps introduced in the last section. This also explains why we apply a recursive approach as in (14) to enforce the time consistency instead of applying directly risk measures to the sum.

Remarks on Assumption (A1)

The Lyapunov-type condition (A1) implies a growth condition with the nonnegative weight function w_0 . Sufficient conditions for (A1) in terms of the risk map \mathcal{R} and the cost function c separately are given as follows:

(A1a) $w_0 : \mathbf{X} \rightarrow [0, \infty)$ is a $\mathcal{B}(\mathbf{X})$ -measurable function satisfying

$$\mathcal{R}_{x,a}(w_0) \vee (-\mathcal{R}_{x,a}(-w_0)) \leq \hat{\gamma}_0 w_0(x) + \hat{K}_0, \forall (x, a) \in \mathbf{K}, \text{ and}$$

(A1b) $|c(x, a)| \leq C_0(w_0^p(x) + 1), \forall (x, a) \in \mathbf{K}$, with some positive constants $p \in (0, 1)$ and $C_0 > 0$.

If, in addition, the risk map is convex, inducing risk-averse behavior, we have that $-\mathcal{R}_{x,a}(-w_0) \leq \mathcal{R}_{x,a}(w_0)$ holds for all $(x, a) \in \mathbf{K}$. Hence, (A1a) reduces to

$$\mathcal{R}_{x,a}(w_0) \leq \hat{\gamma}_0 w_0(x) + \hat{K}_0.$$

Note that for any $\gamma \in (0, 1 - \hat{\gamma}_0)$, there exists a positive constant \tilde{K}_0 such that $C_0(w_0^p + 1) \leq \gamma w_0 + K$. Hence, (A1b) yields $|c(x, a)| \leq \gamma w_0 + K$, which, together with (A1a) implies (A1).

Furthermore, we show below that for coherent risk maps, Assumption (A1) is in fact compatible with the conventional assumptions applied in the literature of MCPs (see e.g., [26, Assumption 10.2.1(d) and (f)] for risk-neutral MCPs).

Assumption 3.4. There exist a $\mathcal{B}(\mathbf{X})$ -measurable function $\tilde{w}_0 : \mathbf{X} \rightarrow [0, \infty)$, positive constants $C, \tilde{\gamma}_0 \in (0, 1)$ and \tilde{K}_0 such that

$$(A1a') \quad \mathcal{R}_{x,a}(\tilde{w}_0) \leq \tilde{\gamma}_0 \tilde{w}_0(x) + \tilde{K}_0, \forall (x, a) \in \mathbf{K}, \text{ and}$$

$$(A1b') \quad |c(x, a)| \leq C(\tilde{w}_0(x) + 1), \forall (x, a) \in \mathbf{K}.$$

PROPOSITION 3.5. *If \mathcal{R} is a coherent risk map, Assumption 3.4 implies (A1) with $w_0 = \frac{C}{\gamma_0 - \tilde{\gamma}_0} \tilde{w}_0$ and $K_0 = C + \frac{C\tilde{K}_0}{\gamma_0 - \tilde{\gamma}_0}$, for any $\gamma_0 \in (\tilde{\gamma}_0, 1)$.*

Proof. Fix $\gamma_0 \in (\tilde{\gamma}_0, 1)$. (A1b') and (A1a') imply for each $(x, a) \in \mathbf{K}$,

$$\begin{aligned} \mathcal{T}_{x,a}(w_0) &= c(x, a) + \mathcal{R}_{x,a}(w_0) \\ &\leq C(1 + \tilde{w}_0(x)) + \mathcal{R}_{x,a}\left(\frac{C}{\gamma_0 - \tilde{\gamma}_0} \tilde{w}_0\right) = C(1 + \tilde{w}_0(x)) + \frac{C}{\gamma_0 - \tilde{\gamma}_0} \mathcal{R}_{x,a}(\tilde{w}_0) \\ &\leq C\left(1 + \frac{\tilde{\gamma}_0}{\gamma_0 - \tilde{\gamma}_0}\right) \tilde{w}_0(x) + K_0 = \gamma_0 w_0(x) + K_0. \end{aligned}$$

Similarly, we obtain $-\mathcal{T}_{x,a}(-w_0) \leq \gamma_0 w_0(x) + K_0$ which implies (A1). \square

3.2 Bounded forward invariant subset

Let $\pi = [\pi_0, \pi_1, \dots]$ be a Markov policy and define the following iteration

$$\mathcal{T}^{(\pi, n)}(v) := \mathcal{T}^{\pi_0}(\mathcal{T}^{\pi_1}(\dots \mathcal{T}^{\pi_n}(v) \dots)), n = 0, 1, \dots$$

Hence, by definition, the n -stage risk-sensitive objective $J_n(x, \pi) = \mathcal{T}_x^{(\pi, n)}(0)$.

Suppose Assumption (A1) holds and $v \in \mathcal{B}_w$ satisfies $|v| \leq w_0 + A$ with some $A > 0$, then Proposition 3.1 shows that $|\mathcal{T}^{(\pi, n)}(v)| \leq w_0 + K(n)$ with some positive $K(n) > 0$. Hence, $\mathcal{T}^{(\pi, n)}(v) \in \mathcal{B}_w, \forall n \in \mathbb{N}$. Note that, however, $\{K(n)\}$ may be an increasing sequence such that $K(n) \rightarrow \infty$ as $n \rightarrow \infty$. This implies that the sequence $\{\mathcal{T}^{(\pi, n)}(v), n = 1, 2, \dots\}$ can be unbounded w.r.t. w -norm. We will specify below a sufficient condition, similar to the ergodicity condition for risk-neutral MCPs (see Assumption 2.2), which implies boundedness of the sequence $\{\mathcal{T}^{(\pi, n)}(v)\}$ w.r.t. the w -seminorm.

Assumption 3.6. Suppose that (A1) holds with some $\mathcal{B}(\mathbf{X})$ -measurable function $w_0 : \mathbf{X} \rightarrow [0, \infty)$, constants $\gamma_0 \in (0, 1)$ and $K_0 > 0$. In addition, assume

(A2) there exists a constant $K > K_0$ such that the inequality

$$\mathcal{R}_{x,a}(w_0 + K) - \mathcal{R}_{x,a}(v) + \mathcal{R}_{y,b}(v) - \mathcal{R}_{y,b}(-w_0 - K) \geq 2K_0$$

holds for all $v \in \mathcal{B}_{1+w_0}$ satisfying $|v| \leq w_0 + K$, and $x, y \in \mathbf{B}_0 := \{x \in \mathbf{X} | w_0(x) \leq \frac{2K_0}{1-\gamma_0}\}$, $a \in \mathbf{A}(x)$, $b \in \mathbf{A}(y)$.

We first state the main result, followed by several remarks on (A2).

THEOREM 3.7. *Suppose Assumption 3.6 holds. Then for each $\pi \in \Delta$,*

$$\|\mathcal{T}^\pi(v)\|_{s, 1+K^{-1}w_0} \leq K, \text{ whenever } \|v\|_{s, 1+K^{-1}w_0} \leq K.$$

Proof. Note that adding a constant to v will not change the required inequality. Due to Lemma 2.1, we may assume that $|v| \leq K + w_0$. By the definition of the weighted seminorm, it is sufficient to show that for each $\pi \in \Delta$,

$$(16) \quad \mathcal{T}_x^\pi(v) - \mathcal{T}_y^\pi(v) \leq 2K + w_0(x) + w_0(y), \forall x \neq y \in \mathbf{X}.$$

We consider the following two cases. Case I: $w_0(x) + w_0(y) \geq \frac{2K_0}{1-\gamma_0}$. This implies

$$(17) \quad 2K + \gamma_0(w_0(x) + w_0(y)) + 2K_0 \leq 2K + w_0(x) + w_0(y).$$

By the monotonicity of \mathcal{R} and Assumption (A1), we have

$$\begin{aligned} \mathcal{T}_x^\pi(v) &\leq \sup_{a \in A(x)} \mathcal{T}_{x,a}(v) \leq \sup_{a \in A(x)} \mathcal{T}_{x,a}(K + w_0) = K + \sup_{a \in A(x)} \mathcal{T}_{x,a}(w_0) \\ &\leq K + \gamma_0 w_0(x) + K_0, \forall x \in \mathbf{X}, \end{aligned}$$

and similarly $\mathcal{T}_y^\pi(v) \geq -K - \gamma_0 w_0(y) - K_0, \forall y \in \mathbf{X}$. Hence, together with (17) we obtain $\mathcal{T}_x^\pi(v) - \mathcal{T}_y^\pi(v) \leq 2K + w_0(x) + w_0(y)$.

Case II: $w_0(x) + w_0(y) \leq \frac{2K_0}{1-\gamma_0}$. Then both x and y are in the subset \mathbf{B}_0 . Hence,

$$\begin{aligned} \mathcal{T}_x^\pi(v) - \mathcal{T}_y^\pi(v) &\leq c^\pi(x) - c^\pi(y) + \mathcal{R}_x^\pi(v) - \mathcal{R}_y^\pi(v) \\ &\text{(by (A2))} \leq 2(K - K_0) + c^\pi(x) - c^\pi(y) + \mathcal{R}_x^\pi(w_0) - \mathcal{R}_y^\pi(w_0) \\ &\text{(by (A1))} \leq 2(K - K_0) + 2K_0 + \gamma_0 w_0(x) + \gamma_0 w_0(y) \\ &\leq 2K + w_0(x) + w_0(y). \end{aligned}$$

Combining I and II, we obtain the required inequality. \square

The above theorem implies immediately the boundedness of iterations $\{\mathcal{T}^{(\pi,n)}(v)\}$ under the weighted seminorm with $w = 1 + K^{-1}w_0$.

COROLLARY 3.8. *Suppose Assumption 3.6 holds. Then for each $\pi \in \Pi_M$ and $n \in \mathbb{N}$, $\|\mathcal{T}^{(\pi,n)}(v)\|_{s,1+K^{-1}w_0} \leq K$, whenever $\|v\|_{s,1+K^{-1}w_0} \leq K$.*

Hence, given a risk map satisfying (A1) and (A2), we can restrict ourselves to the bounded forward invariant subset $\mathcal{B}_w^{(K)}$ (with $w = 1 + K^{-1}w_0$) defined as

$$\mathcal{B}_w^{(K)} := \{v \in \mathcal{B}_w \mid \|v\|_{s,w} \leq K\},$$

rather than the whole set \mathcal{B}_w .

Remarks on (A2)

(A2) in fact generalizes Doeblin condition stated in Assumption 2.2(ii). To see this, taking the risk-neutral risk map $\mathcal{R}_{x,a}(\cdot) = Q_{x,a}[\cdot]$, (A2) becomes $Q_{x,a}[w_0 + K - v] - Q_{y,b}[-v - w_0 - K] \geq 2K_0$ due to the linearity of Q . Suppose now Q satisfies Assumption 2.2(ii) on \mathbf{B}_0 , i.e., there exists a probability measure μ and some constant $\alpha \in (0, 1)$ such that $Q_{x,a}(\mathbf{B}) \geq \alpha\mu(\mathbf{B}), \forall \mathbf{B} \in \mathcal{B}(\mathbf{X}), x \in \mathbf{B}_0$ and $a \in A(x)$. Then, it is easy to see that

$$Q_{x,a}[w_0 + K - v] - Q_{y,b}[-v - w_0 - K] \geq 2\alpha\mu[w_0 + K] \geq 2\alpha K.$$

Hence, given K_0 and α , we can easily choose K such that (A2) holds. This connection holds also for coherent risk maps, which we will show below. For more general risk maps, however, the connection is not straightforward. We will show the sufficient conditions for (A2) in the case of entropic maps in Section 4 and some classes of utility-based shortfalls in Section 5.4.

Conditions for coherent risk maps

For a special case, when applying coherent risk maps, we show below that the above set of conditions can be replaced by a more conventional set of conditions.

Assumption 3.9. Suppose (A1b') and (A1a') hold with some nonnegative $\mathcal{B}(\mathbf{X})$ -measurable function \tilde{w}_0 , positive constants C , $\tilde{\gamma}_0 \in (0, 1)$ and \tilde{K}_0 . In addition,

(A2') there exist a probability measure $\mu \in \mathcal{P}_{1+\tilde{w}_0}$, a constant $\alpha > 0$ such that

$$\mathcal{R}_{x,a}(v) - \mathcal{R}_{x,a}(u) \geq \alpha\mu[v - u], \forall v \geq u \in \mathcal{B}_{1+\tilde{w}_0}$$

holds for all $x \in \mathbf{B}_0 := \{x \in \mathbf{X} | \tilde{w}_0(x) \leq 2R_0\}$ and $a \in \mathbf{A}(x)$, with some $R_0 > \tilde{K}_0/(1 - \tilde{\gamma}_0)$.

PROPOSITION 3.10. *Let \mathcal{R} be a coherent risk map w.r.t. an MCP. Suppose Assumption 3.9 holds. Then $\gamma_0 := \frac{R_0 - \tilde{K}_0 + \tilde{\gamma}_0}{1 + R_0} \in (\tilde{\gamma}_0, 1)$ and Assumption 3.6 holds with γ_0 , $w_0 = \frac{C}{\gamma_0 - \tilde{\gamma}_0} \tilde{w}_0$, $K_0 = C + \frac{C\tilde{K}_0}{\gamma_0 - \tilde{\gamma}_0}$ and $K = K_0/\alpha$.*

Proof. First, the inequalities $\tilde{\gamma}_0 < \gamma_0 < 1$ can be easily verified and therefore the proof is omitted here. Second, Proposition 3.5 implies (A1).

It remains to verify (A2'). First, it is easy to verify that $\frac{CR_0}{\gamma_0 - \tilde{\gamma}_0} = \frac{K_0}{1 - \gamma_0}$ and

$$\mathbf{B}_0 = \{x \in \mathbf{X} | \tilde{w}_0(x) \leq 2R_0\} = \left\{x \in \mathbf{X} \left| w_0(x) \leq \frac{2K_0}{1 - \gamma_0} \right.\right\}.$$

Let $w' := w_0 + K$. Then for any $v \in \mathcal{B}_w$ satisfying $|v| \leq w'$ and $x \in \mathbf{B}_0, a \in \mathbf{A}(x)$, (A2') implies $\mathcal{R}_{x,a}(v) - \mathcal{R}_{x,a}(w') \leq \alpha\mu(v - w')$ and hence,

$$\begin{aligned} \mathcal{R}_{x,a}(v) - \alpha\mu(v) &\leq \mathcal{R}_{x,a}(w') - \alpha\mu(w') \\ &= (1 - \alpha)K + \mathcal{R}_{x,a}(w_0) - \alpha\mu(w_0) \leq (1 - \alpha)K + \mathcal{R}_{x,a}(w_0). \end{aligned}$$

Repeating the above derivation for $-v \leq w'$, we obtain $\mathcal{R}_{x,a}(-v) - \alpha\mu(-v) \leq (1 - \alpha)K + \mathcal{R}_{x,a}(w_0)$. The coherency of \mathcal{R} yields $\mathcal{R}_{x,a}(-v) \geq -\mathcal{R}_{x,a}(v)$ and hence,

$$-\mathcal{R}_{x,a}(v) + \alpha\mu(v) \leq (1 - \alpha)K + \mathcal{R}_{x,a}(w_0).$$

Thus, for any $x, y \in \mathbf{B}_0, a \in \mathbf{A}(x), b \in \mathbf{A}(y)$, we have

$$\mathcal{R}_{x,a}(v) - \mathcal{R}_{y,b}(v) \leq 2(1 - \alpha)K + \mathcal{R}_{x,a}(w_0) + \mathcal{R}_{y,b}(w_0).$$

This implies that (A2) holds with $K = K_0/\alpha$. □

3.3 Geometric contraction

In this subsection, we state sufficient conditions, under which \mathcal{R}^π is a contraction, similar to the geometric contraction for standard Markov chains stated in (5). We first introduce the concept of *upper envelope*.

DEFINITION 3.11. A coherent risk map $\bar{\mathcal{R}}^{(w,K)}$ is said to be an upper envelope of a valuation map \mathcal{R} given a bound $K \in \mathbb{R}_+$, if for all $v, u \in \mathcal{B}_w^{(K)}$,

$$\mathcal{R}_{x,a}(v) - \mathcal{R}_{x,a}(u) \leq \bar{\mathcal{R}}_{x,a}^{(w,K)}(v - u), \forall (x, a) \in \mathbf{K}.$$

Let w_0 be a function satisfying Assumption (A1) and $w = 1 + K^{-1}w_0$ with K satisfying (A2). Analogous to the assumption applied to risk-neutral MCPs (see Assumption 2.2) and the one applied in risk-sensitive MCPs by Shen et al. [45], we introduce a set of conditions on the upper envelope.

Assumption 3.12. (B1) there exist constants $\gamma \in (0, 1)$, $\bar{K} > 0$ and an upper envelope $\bar{\mathcal{R}}^{(w,K)}$ such that

$$\bar{\mathcal{R}}_{x,a}^{(w,K)}(w_0) \leq \gamma w_0(x) + \bar{K}, \forall (x, a) \in \mathbf{K}.$$

(B2) there exist a constant $\alpha \in (0, 1)$ and a probability measure $\mu \in \mathcal{P}_{1+w_0}$ satisfying

$$\bar{\mathcal{R}}_{x,a}^{(w,K)}(v) - \bar{\mathcal{R}}_{x,a}^{(w,K)}(u) \geq \alpha \mu[v - u], \forall x \in \mathbf{B}, a \in \mathbf{A}(x), v \geq u \in \mathcal{B}_{1+w_0}$$

where $\mathbf{B} := \{x \in \mathbf{X} | w_0(x) \leq R\}$ for some $R > \frac{2\bar{K}}{1-\gamma}$.

Remark 3.13. Apparently, if \mathcal{R} is coherent, then \mathcal{R} is an upper envelope of itself for all bounds $K > 0$, due to its subadditivity stated in (6). Hence, Assumption 3.12 is equivalent to Assumptions (A1a') and (A2'). On the other hand, we have shown in Proposition 3.10 that Assumptions (A1b'), (A1a') and (A2') imply Assumption 3.6. Hence, Assumptions 3.6 and 3.12 can be reduced to Assumption 3.9.

We state now the main result of this subsection.

THEOREM 3.14. Suppose Assumption 3.12 holds. Then there exists a constant $\bar{\alpha} \in (0, 1)$ and $\beta > 0$ such that

$$\|\mathcal{R}^\pi(v) - \mathcal{R}^\pi(u)\|_{s, 1+\beta w_0} \leq \bar{\alpha} \|v - u\|_{s, 1+\beta w_0}, \forall v, u \in \mathcal{B}_w^{(K)}, \pi \in \Delta.$$

Proof. The proof is in essence similar to the proof of Shen et al. [45, Theorem 3.11]. For readers' convenience, we incorporate it in the Appendix. \square

The Lyapunov-type condition (B1) yields the following result.

LEMMA 3.15. Suppose that Assumptions 3.6 and (B1) hold. Then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \|\mathcal{T}^{(\pi,n)}(v) - \mathcal{T}^{(\pi,n)}(u)\|_w = 0, \forall v, u \in \mathcal{B}_w^{(K)}, \pi \in \Pi_M.$$

Proof. It is sufficient to show that $\|\mathcal{T}^{(\pi,n)}(v) - \mathcal{T}^{(\pi,n)}(u)\|_w$ is uniformly bounded. Indeed, let $\mathcal{U}_x^\pi(v) := \int \pi(da|x) \bar{\mathcal{R}}_{x,a}^{(w,K)}(v)$ and $K' := \frac{\bar{K}}{K} + 1 - \gamma > 0$. By (B1), we have for each $x \in \mathbf{X}$ and $\pi \in \Delta$,

$$\mathcal{T}_x^\pi(v) - \mathcal{T}_x^\pi(u) \leq \mathcal{U}_x^\pi(|v - u|) \leq \|v - u\|_w \left(\frac{\gamma}{K} w_0(x) + \frac{\bar{K}}{K} + 1 \right),$$

which implies that $|\mathcal{T}_x^\pi(v) - \mathcal{T}_x^\pi(u)| \leq \|v - u\|_w(\gamma w(x) + K')$.

In addition, by Theorem 3.7, $\|\mathcal{T}^{(\pi,n)}(v)\|_{s,w} \leq K$ holds for all $n \in \mathbb{N}$. Hence, by induction w.r.t. n , we have for $n = 2, 3, \dots$

$$\begin{aligned} |\mathcal{T}_x^{(\pi,n)}(v) - \mathcal{T}_x^{(\pi,n)}(u)| &\leq \mathcal{U}_x^{\pi_0}(|\mathcal{T}^{([\pi_1, \dots], n-1)}(v) - \mathcal{T}^{([\pi_1, \dots], n-1)}(u)|) \\ &\leq \|v - u\|_w \mathcal{M}_x^{\pi_0} \left(\gamma^{n-1} w + K' \sum_{k=0}^{n-2} \gamma^k \right) \\ &\leq \|v - u\|_w \left(\gamma^n w(x) + K' \sum_{k=0}^{n-1} \gamma^k \right), \forall x \in \mathbf{X}, \end{aligned}$$

which implies that $\|\mathcal{T}^{(\pi,n)}(v) - \mathcal{T}^{(\pi,n)}(u)\|_w \leq \frac{K'}{1-\gamma}$. \square

3.4 Nonlinear Poisson equation

Similar to the risk-neutral average MCPs (see, e.g., [26, Chapter 10]), the optimization problem of average risk-sensitive objective defined in (15) is closely related to the following nonlinear Poisson equation

$$(18) \quad \rho + h(x) = \inf_{a \in \mathbf{A}(x)} (c(x, a) + \mathcal{R}(h|x, a)), \forall x \in \mathbf{X},$$

where $(\rho, h) \in \mathbb{R} \times \mathcal{B}_w$ is its solution. Define the following operator

$$(19) \quad \mathcal{F}_x(v) = \mathcal{F}(v|x) := \inf_{a \in \mathbf{A}(x)} (c(x, a) + \mathcal{R}(v|x, a)).$$

Existence of optimal selectors

In order to guarantee the existence of an “optimal selector” and the measurability of the operator \mathcal{F} , we assume

Assumption 3.16. For all $x \in \mathbf{X}$,

- (C1) the action space $\mathbf{A}(x)$ is compact;
- (C2) the cost function $c(x, a)$ is lower semicontinuous (l.s.c.) on $\mathbf{A}(x)$;
- (C3) $a \mapsto \mathcal{R}(v|x, a)$ is l.s.c. on $\mathbf{A}(x)$, for each $v \in \mathcal{B}_w$ satisfying $|v| \leq K + w_0$.

Note that each $v \in \mathcal{B}_w^{(K)}$ can be represented as $v = \tilde{v} + C$ with some constant C and $\tilde{v} \in \mathcal{B}_w$ satisfying $|\tilde{v}| \leq K + w_0$. Due to the translation invariance, (C3) immediately imply that $a \mapsto \mathcal{R}(v|x, a)$ is l.s.c. on $\mathbf{A}(x)$, for each $v \in \mathcal{B}_w^{(K)}$ and $x \in \mathbf{X}$. Hence, a direct application of [26, Lemma 8.3.8(a)] yields

PROPOSITION 3.17. *Let \mathcal{R} be a risk map on an MCP. Suppose Assumption 3.16 holds. Then for each $v \in \mathcal{B}_w^{(K)}$ and $x \in \mathbf{X}$, there exists a deterministic policy $f \in \Delta_D$ such that*

$$c^f(x) + \mathcal{R}^f(v|x) = \mathcal{F}(v|x) := \inf_{a \in \mathbf{A}(x)} (c(x, a) + \mathcal{R}(v|x, a))$$

and furthermore, $\mathcal{F}(v|\cdot)$ is $\mathcal{B}(\mathbf{X})$ -measurable.

Remark 3.18. In risk-neutral MCPs, where $\mathcal{R}_{x,a}(\cdot) = Q_{x,a}[\cdot]$, (C3) is satisfied (see e.g., [26, Lemma 8.3.7(a)]) if

(C3a) $a \mapsto \mathcal{R}_{x,a}[v]$ is l.s.c. on $A(x)$, for each $v \in \mathcal{B}$; and

(C3b) $a \mapsto \mathcal{R}_{x,a}[w_0]$ is l.s.c. on $A(x)$.

For general risk maps, we need to further assume that \mathcal{R} is l.s.c. w.r.t. v on $\mathcal{B}_w^{(K)}$:

(C3c) Let $\{v_n \in \mathcal{B}, n = 0, 1, \dots\}$ be a sequence of functions that pointwise converges to $v \in \mathcal{B}_w^{(K)}$. Then for all $(x, a) \in K$, $\liminf_{n \rightarrow \infty} \mathcal{R}(v_n|x, a) \geq \mathcal{R}(v|x, a)$.

By extending the proof of [26, Lemma 8.3.7(a)], (C3a) – (C3c) imply (C3). Note that in the case of risk-neutral MCPs, (C3c) is already satisfied due to Fatou's lemma. However, it does not hold for general risk maps (see [12] for a detailed discussion). Note that because verification of (C3c) is usually not straightforward, we will verify (C3) directly for various examples of risk maps shown in Section 5.

Existence of a unique solution

We show now the operator \mathcal{F} is also a contraction under the weighted seminorm.

LEMMA 3.19. *Suppose Assumptions 3.12 and 3.16 hold. Then there exists a constant $\bar{\alpha} \in (0, 1)$ and $\beta > 0$ such that*

$$\|\mathcal{F}(v) - \mathcal{F}(u)\|_{s, 1+\beta w_0} \leq \bar{\alpha} \|v - u\|_{s, 1+\beta w_0}, \forall v, u \in \mathcal{B}_w^{(K)}.$$

Proof. By definition we have

$$\mathcal{F}_x(v) - \mathcal{F}_x(u) \leq \sup_{a \in A(x)} \{\mathcal{R}_{x,a}(v) - \mathcal{R}_{x,a}(u)\} \leq \sup_{a \in A(x)} \bar{\mathcal{R}}_{x,a}^{(w,K)}(v - u), \forall x \in X.$$

Define $\mathcal{U}_x(v) := \sup_{a \in A(x)} \bar{\mathcal{R}}_{x,a}^{(w,K)}(v)$ and we have

$$\mathcal{F}_x(v) - \mathcal{F}_x(u) \leq \mathcal{U}_x(v - u) \leq \mathcal{U}_x(|v - u|).$$

The rest of the proof is similar to the proof of Theorem 3.14. \square

THEOREM 3.20. *Suppose that Assumptions 3.6, 3.12 and 3.16 hold. Then there exist a unique $\rho \in \mathbb{R}$ and $h \in \mathcal{B}_w$ such that (ρ, h) satisfies the Poisson equation*

$$(20) \quad \rho + h(x) = \inf_{a \in A(x)} (c(x, a) + \mathcal{R}(h|x, a)), \forall x \in X.$$

In particular, if \mathcal{R} is coherent, the assertion holds under Assumptions 3.9 and 3.16.

Proof. Let $\hat{w} := 1 + \beta w_0$ as in Lemma 3.19. Then the map $\mathcal{F} : \mathcal{B}_w^{(K)} \rightarrow \mathcal{B}_w^{(K)}$ defined in 19 is a contraction under \hat{w} -seminorm. Note that $\mathcal{B}_w^{(K)} \subset \mathcal{B}_{\hat{w}}$. We now extend the fixed-point theorem w.r.t. span-seminorm (cf. p. 321 [1] for bounded \hat{w}) to \hat{w} -seminorm. Let $\tilde{\mathcal{B}}_{\hat{w}} = \mathcal{B}_{\hat{w}} / \sim$ be the quotient space, which is induced

by the equivalence relation \sim on $\mathcal{B}_{\hat{w}}$ defined by $v \sim u$ if and only if there exists some constant $C \in \mathbb{R}$ such that $v(x) - u(x) = C$ for all $x \in \mathbf{X}$, endowed with the quotient norm induced by the \hat{w} -seminorm. For $v \in \mathcal{B}_{\hat{w}}$, let \tilde{v} be the corresponding equivalence class in $\tilde{\mathcal{B}}_{\hat{w}}$ and $\tilde{\mathcal{F}} : \tilde{\mathcal{B}}_{\hat{w}} \rightarrow \tilde{\mathcal{B}}_{\hat{w}}$ be the canonically induced map, i.e., $\tilde{\mathcal{F}}(\tilde{v}) := \widetilde{\mathcal{F}(v)}$, $v \in \mathcal{B}_{\hat{w}}$. Since \mathcal{F} is a contraction w.r.t. \hat{w} -seminorm on $\mathcal{B}_w^{(K)} \subset \mathcal{B}_{\hat{w}}$, $\tilde{\mathcal{F}}$ is a contraction on $\{v \in \tilde{\mathcal{B}}_{\hat{w}} \mid \|v\|_{s,w} \leq K\}$ and therefore has a unique fixed point. Conversely, it follows that the map \mathcal{F} has a \hat{w} -seminorm fixed point. In other words, there exists $h \in \mathcal{B}_w^{(K)}$ such that $\|\mathcal{F}(h) - h\|_{s,w} = 0$ and $\rho := \mathcal{F}(h) - h$ is a constant.

Next we show that such ρ is unique. Suppose there exists another solution $(\rho', h') \in \mathbb{R} \times \mathcal{B}_w^{(K)}$ to the Poisson equation. Then $\mathcal{F}^n(h') = n\rho' + h'$ and $\mathcal{F}^n(h) = n\rho + h$. However, by Lemma 3.15, we have

$$\frac{1}{n} \|n\rho' + h' - n\rho - h\|_w = \frac{1}{n} \|\mathcal{F}^n(h') - \mathcal{F}^n(h)\|_w \rightarrow 0,$$

which implies $\rho' = \rho$. \square

3.5 Solution to average risk-sensitive MCPs

We show below that the unique real value ρ satisfying the Poisson equation is in fact *optimal* for the average risk-sensitive objective defined in (15).

THEOREM 3.21. *Suppose that Assumptions 3.6, 3.12 and 3.16 hold. Let (ρ, h) be a solution to the Poisson equation defined in (18). Then $\rho = J^*(x) = J(x, f^\infty)$, $\forall x \in \mathbf{X}$, where f denotes the optimal selector in the right-hand side of the Poisson equation. In particular, if \mathcal{R} is coherent, the assertion holds under Assumptions 3.9 and 3.16.*

Proof. Let \mathcal{T}^π be the operator defined in (13). By its definition, $\mathcal{F}(h) = \mathcal{T}^f(h)$ and the induction $(\mathcal{T}^f)^n(h) = (\mathcal{T}^f)^{n-1}(\rho + h) = n\rho + h$, $n = 1, 2, \dots$, yields $\frac{1}{n} \|(\mathcal{T}^f)^n(h) - \rho\|_w \rightarrow 0$ as $n \rightarrow \infty$. On the other hand, for any $v \in \mathcal{B}_w^{(K)}$, by Lemma 3.15, $\frac{1}{n} \|(\mathcal{T}^f)^n(v) - (\mathcal{T}^f)^n(h)\|_w \rightarrow 0$ implies that

$$J(x, f^\infty) = \lim_{n \rightarrow \infty} \frac{1}{n} (\mathcal{T}^f)^n_x(0) = \rho, \forall x \in \mathbf{X}.$$

Next we prove that $\rho \leq J(x, \pi)$ for all $\pi \in \Pi_M = \Delta^\infty$ and $x \in \mathbf{X}$. In fact, let $\pi = [\pi_0, \pi_1, \dots]$ be an arbitrary Markov policy. By definition $h \leq \mathcal{T}^\pi(h) - \rho$, $\forall \pi \in \Delta$. Iterating this inequality yields $h \leq \mathcal{T}^{\pi_0}(\mathcal{T}^{\pi_1}(\dots \mathcal{T}^{\pi_{n-1}}(h) \dots)) - n\rho$, and hence

$$(21) \quad 0 \leq \limsup_{n \rightarrow \infty} \frac{1}{n} \mathcal{T}^{\pi_0}(\mathcal{T}^{\pi_1}(\dots \mathcal{T}^{\pi_{n-1}}(h) \dots)) - \rho.$$

Note that by definition $J(x, \pi) = \limsup_{n \rightarrow \infty} \frac{1}{n} \mathcal{T}_x^{\pi_0}(\mathcal{T}^{\pi_1}(\dots \mathcal{T}^{\pi_{n-1}}(0) \dots))$. Lemma 3.15 yields $J(x, \pi) = \limsup_{n \rightarrow \infty} \frac{1}{n} \mathcal{T}_x^{\pi_0}(\mathcal{T}^{\pi_1}(\dots \mathcal{T}^{\pi_{n-1}}(h) \dots))$. Hence, (21) implies

$$\rho \leq \inf_{\pi \in \Pi_M} J(\pi) = J^*.$$

Since f^∞ is a valid Markov policy in Π_M , $\rho = J^*(x) = J(x, f^\infty)$, $\forall x \in \mathbf{X}$. \square

Value iteration We can use the following iterative procedure to calculate the optimal average risk: start from any $v_0 \in \mathcal{B}_w^{(K)}$ and iterate

$$v_{n+1} = \mathcal{F}(v_n) \text{ with } f_n \text{ being its selector, } n = 0, 1, 2, \dots$$

Lemma 3.19 and Theorem 3.20 ensure that $v_n \rightarrow \rho$ and $f_n \rightarrow f$, as $n \rightarrow \infty$, where $\rho = J^*$ is the unique solution to the Poisson equation and f the optimal policy. In particular, this iteration will geometrically converge with rate $\bar{\alpha}$ as in Lemma 3.19.

4 Entropic maps

In this section, we investigate the sufficient conditions particularly for the entropic map, which is the most widely used risk measure in the last four decades. Recall that in Example 2.7, the entropic map is defined as

$$\mathcal{R}_{x,a}(v) := \frac{1}{\lambda} \ln(Q_{x,a}[e^{\lambda v}]) = \frac{1}{\lambda} \ln \left\{ \int_{\mathbf{X}} Q(dy|x, a) e^{\lambda v(y)} \right\}.$$

We consider in this section mainly the case $\lambda > 0$ and hence \mathcal{R} is convex, which induces risk-averse behavior, though the results stated below can be easily extended to negative λ . From now on, without loss of generality, we set $\lambda = 1$.

4.1 Upper envelope

We first derive its upper envelope.

PROPOSITION 4.1. *Let μ be a probability measure on $(\mathbf{X}, \mathcal{B}(\mathbf{X}))$ and $\nu(v) := \ln(\mu[e^v])$. Suppose $\mu[e^v] < \infty$ holds for all $v \in \mathcal{B}_w$. Then*

$$(i) \quad \nu(v) \leq \frac{\mu[e^v v]}{\mu[e^v]}, \text{ and}$$

$$(ii) \quad \nu(v) - \nu(u) \leq \sup_{f \in \mathcal{B}_w^{(K)}} \frac{\mu[e^f(v-u)]}{\mu[e^f]}, \quad \forall v, u \in \mathcal{B}_w^{(K)}.$$

Proof. Given any two $u, v \in \mathcal{B}_w$, we obtain

$$(22) \quad \nu(v) - \nu(u) = \ln \frac{\mu[e^v]}{\mu[e^u]} = \ln \frac{\mu[e^u e^{v-u}]}{\mu[e^u]} \geq \frac{\mu[e^u(v-u)]}{\mu[e^u]},$$

where the last inequality is due to Jensen's inequality. Hence,

$$\ln(\mu[e^v]) \geq \frac{\mu[e^u v]}{\mu[e^u]} - \mu \left[\frac{e^u}{\mu[e^u]} (u - \ln(\mu[e^u])) \right], \quad \forall u, v \in \mathcal{B}_w.$$

Restricting u and v to be in the subset $\mathcal{B}_w^{(K)}$, the above inequality yields

$$\ln(\mu[e^v]) \geq \sup_{\xi = \frac{e^u}{\mu[e^u]}, u \in \mathcal{B}_w^{(K)}} \mu[\xi v] - \mu[\xi \ln(\xi)].$$

Since the equality holds by taking $\xi^* := \frac{e^v}{\mu[e^v]}$, we obtain

$$(23) \quad \ln(\mu[e^v]) = \sup_{\xi = \frac{e^u}{\mu[e^u]}, u \in \mathcal{B}_w^{(K)}} \mu[\xi v] - \mu[\xi \ln(\xi)].$$

The second term $\mu[\xi \ln(\xi)]$ on the right-hand side of the above equation is the *relative entropy* and is always nonnegative (for proof see, e.g., [33, Section 5.1]). Hence, we obtain (i). Finally, (ii) follows from

$$\ln(\mu[e^v]) - \ln(\mu[e^u]) \leq \sup_{\xi = \frac{e^f}{\mu[e^f]}, f \in \mathcal{B}_w^{(K)}} \mu[\xi(v - u)] = \sup_{f \in \mathcal{B}_w^{(K)}} \frac{\mu[e^f(v - u)]}{\mu[e^f]}.$$

□

Remark 4.2. The inequality in (23) is similar to the dual representation of convex risk measures on L^∞ [19, 20] or on more general spaces such as *Orlicz hearts* [8]. However, since we consider a different function space, i.e., the weighted norm space, the existing result cannot be directly applied here. On the other hand, for other types of convex valuation functions, their dual representation provide us a generic approach to calculate their upper envelopes, as shown in the above proposition.

By Proposition 4.1, we obtain one upper envelope for the entropic map:

$$(24) \quad \bar{\mathcal{R}}_{x,a}^{(w,K)}(u) = \sup_{f \in \mathcal{B}_w^{(K)}} \frac{Q_{x,a}[e^f u]}{Q_{x,a}[e^f]},$$

provided that $Q_{x,a}[e^f] < \infty$ holds for all $f \in \mathcal{B}_w$ and $(x, a) \in \mathbf{K}$. We show below how this condition is satisfied.

Assumption 4.3. There exist a $\mathcal{B}(\mathbf{X})$ -measurable function $w_1 : \mathbf{X} \rightarrow [1, \infty)$, positive constants $\gamma_1 \in (0, 1)$ and $K_1 > 0$ such that $\mathcal{R}_{x,a}(w_1) \leq \gamma_1 w_1(x) + K_1$.

If the above assumption holds and setting $w_0 := w_1^p$ with any $p \in (0, 1)$, then for all $f \in \mathcal{B}_{w_0} \subseteq \mathcal{B}_w$, there exists a constant K_f (depending on p and $\|f\|_{w_0}$) satisfying $|f(x)| \leq \|f\|_{w_0} w_0(x) \leq w_1(x) + K_f, \forall x \in \mathbf{X}$. We immediately have

$$Q_{x,a}[e^f] \leq Q_{x,a}[e^{w_1 + K_f}] \leq e^{K_f} e^{\gamma_1 w_1 + K_1} < \infty, \forall (x, a) \in \mathbf{K}$$

and therefore, the upper envelope for the entropic map in (24) is well-defined. In the following theorem, we show that $w_0 = w_1^p$ with any $p \in (0, 1)$ satisfies (B1) for the upper envelope of the entropic map with some constants $\gamma_2 \in (0, 1)$ and $K_2 > 0$.

THEOREM 4.4. *Suppose that Assumption 4.3 holds. Let $w_0 := w_1^p$ with $p \in (0, 1)$. Then, for any constant $K > 0$, there exist constants $\gamma_2 \in (0, 1)$ (depending only on p and γ_1) and $K_2 > 0$ (depending on p, K, λ_1 and K_1) such that*

$$\sup_{f \in \mathcal{B}_{w_0}^{(K)}} \frac{Q_{x,a}[e^f w_0]}{Q_{x,a}[e^f]} \leq \gamma_2 w_0(x) + K_2, \forall (x, a) \in \mathbf{K}.$$

Since the proof is rather technical, we postpone it to the Appendix.

Note that by the definition of upper envelope, $\mathcal{R}_{x,a}(w_0) \leq \sup_{f \in \mathcal{B}_{w_0}^{(K)}} \frac{Q_{x,a}[e^f w_0]}{Q_{x,a}[e^f]}$. Hence, the above theorem implies immediately the following corollary.

COROLLARY 4.5. *Suppose that Assumption 4.3 holds. Then, for any $p \in (0, 1)$, $w_0 := w_1^p$, there exist constants $\hat{\gamma}_0 \in (0, 1)$ (depending on p and γ_1) and \hat{K}_0 (depending on p , γ_1 and K_1) satisfying $\mathcal{R}_{x,a}(w_0) \leq \hat{\gamma}_0 w_0(x) + \hat{K}_0, \forall (x, a) \in \mathbf{K}$.*

In summary, if Assumption 4.3 holds, then

1. by Corollary 4.5, Assumption (A1a') stated in Section 3.1 holds with w_0 and constants $\hat{\gamma}_0$ and \hat{K}_0 , and if in addition, the cost function c satisfies $|c| \leq \tilde{\gamma}_0 w_0 + C_0$ with some $\tilde{\gamma}_0 \in (0, 1 - \gamma_0)$ and $C_0 > 0$, then (A1) holds;
2. by Theorem 4.4, Assumption (B1) stated in Section 3.3 holds with the same w_0 and constants γ_2 and K_2 .

4.2 Doeblin-type conditions

We introduce the following notation of *level-sets*. For any unbounded nonnegative $\mathcal{B}(\mathbf{X})$ -measurable function w and any real number $R \in \mathbb{R}$, we define $\mathbf{B}_w(R) := \{x \in \mathbf{X} | w(x) \leq R\}$ and $\mathbf{B}_w^c(R)$ its complementary set. We investigate in this section the properties of the entropic map restricted to bounded level-sets. We first introduce the *local Doeblin condition* (see [16] and references therein) as follows.

Assumption 4.6. Let $w_0 : \mathbf{X} \rightarrow [0, \infty)$ be a $\mathcal{B}(\mathbf{X})$ -measurable function. For any level-set $\mathbf{C} := \mathbf{B}_{w_0}(R)$, $R > 0$, there exist a measure $\mu_{\mathbf{C}}$ and constants $\lambda_{\mathbf{C}}^+, \lambda_{\mathbf{C}}^- > 0$ such that $\mu_{\mathbf{C}}(\mathbf{C}) > 0$ and

$$\lambda_{\mathbf{C}}^- \mu_{\mathbf{C}}(\mathbf{D} \cap \mathbf{C}) \leq Q_{x,a}(\mathbf{D} \cap \mathbf{C}) \leq \lambda_{\mathbf{C}}^+ \mu_{\mathbf{C}}(\mathbf{D} \cap \mathbf{C}), \forall x \in \mathbf{C}, a \in \mathbf{A}(x), \forall \mathbf{D} \in \mathcal{B}(\mathbf{X}).$$

Remark 4.7. This assumption is stronger than the standard Doeblin condition (see Assumption 2.2(ii)). In fact, it is easy to verify that the following two conditions are equivalent:

- (i) There exist a measure $\mu_{\mathbf{C}}$ and a constant $\lambda_{\mathbf{C}}^- > 0$ such that $\mu_{\mathbf{C}}(\mathbf{C}) > 0$ and

$$(25) \quad Q_{x,a}(\mathbf{D} \cap \mathbf{C}) \geq \lambda_{\mathbf{C}}^- \mu_{\mathbf{C}}(\mathbf{D} \cap \mathbf{C}), \forall x \in \mathbf{C}, a \in \mathbf{A}(x), \mathbf{D} \in \mathcal{B}(\mathbf{X}).$$

- (ii) There exist a probability measure μ and a constant $\alpha > 0$ such that

$$(26) \quad Q_{x,a}(\mathbf{D}) \geq \alpha \mu(\mathbf{D}), \forall x \in \mathbf{C}, a \in \mathbf{A}(x), \mathbf{D} \in \mathcal{B}(\mathbf{X}).$$

THEOREM 4.8. *Suppose Assumption 4.6 and Assumption 4.3 hold. Let w_1 be the weight function as in Assumption 4.3 and $\mathbf{B} = \mathbf{B}_{w_0}(R_0)$ be a bounded level-set w.r.t. some $R_0 > 0$, where $w_0 := w_1^p$, $p \in (0, 1)$. Then for any positive constant $K_0 > 0$, there exists a positive constant $K > K_0$ such that*

$$\mathcal{R}_{x,a}(v) - \mathcal{R}_{y,b}(v) \leq 2(K - K_0) + \mathcal{R}_{x,a}(w_0) - \mathcal{R}_{y,b}(-w_0)$$

holds for all $x, y \in \mathbf{B}$, $a \in \mathbf{A}(x)$, $b \in \mathbf{A}(y)$ and $v \in \mathcal{B}_{1+w_0}$ satisfying $|v| \leq w_0 + K$.

Proof. Let $C := B_{w_0}(R) \supset B = B_{w_0}(R_0)$ with $R > R_0$. Let $\mathbf{1}_D(\cdot)$ be the indicator function on X for any $D \subset X$. Then

$$(27) \quad \frac{Q_{x,a}[e^v]}{Q_{y,b}[e^v]} = \frac{Q_{x,a}[e^v \mathbf{1}_C] + Q_{x,a}[e^v \mathbf{1}_{C^c}]}{Q_{y,b}[e^v \mathbf{1}_C] + Q_{y,b}[e^v \mathbf{1}_{C^c}]} \leq \frac{Q_{x,a}[e^v \mathbf{1}_C] + Q_{x,a}[e^v \mathbf{1}_{C^c}]}{Q_{y,b}[e^v \mathbf{1}_C]}.$$

We first consider the second quotient. By $|v| \leq K + w_0$, we obtain

$$\frac{Q_{x,a}[e^v \mathbf{1}_{C^c}]}{Q_{y,b}[e^v \mathbf{1}_C]} \leq e^{2K} \frac{Q_{x,a}[e^{w_0} \mathbf{1}_{C^c}]}{Q_{y,b}[e^{-w_0} \mathbf{1}_C]} = e^{2K} \frac{\theta(x, a, C) Q_{x,a}[e^{w_0}]}{\theta'(y, b, C) Q_{y,b}[e^{-w_0}]}$$

where we define $\theta(x, a, C) := \frac{Q_{x,a}[e^{w_0} \mathbf{1}_{C^c}]}{Q_{x,a}[e^{w_0}]}$ and $\theta'(y, b, C) := \frac{Q_{y,b}[e^{-w_0} \mathbf{1}_C]}{Q_{y,b}[e^{-w_0}]}$. By Theorem 4.4, there exist some constants $\gamma_2 \in (0, 1)$ and $K_2 > 0$ such that

$$\begin{aligned} \theta(x, a, C) &\leq \|\mathbf{1}_{C^c}\|_{w_0} \frac{Q_{x,a}[e^{w_0} w_0]}{Q_{x,a}[e^{w_0}]} \leq \|\mathbf{1}_{C^c}\|_{w_0} \sup_{|v| \leq w_0} \frac{Q_{x,a}[e^v w_0]}{Q_{x,a}[e^v]} \\ &\leq \|\mathbf{1}_{C^c}\|_{w_0} (\gamma_2 w_0(x) + K_2). \end{aligned}$$

Hence, $\theta(x, a, C) \leq \|\mathbf{1}_{C^c}\|_{w_0} \sup_{x \in B} (\gamma_2 w_0(x) + K_2) \frac{\gamma_2 R_0 + K_2}{R}$. Similarly,

$$\theta'(y, b, C) = 1 - \frac{Q_{y,b}[e^{-w_0} \mathbf{1}_{C^c}]}{Q_{y,b}[e^{-w_0}]} \geq 1 - \frac{\gamma_2 R_0 + K_2}{R}.$$

Hence, $\sup_{x,y \in B, a \in A(x), b \in A(y)} \frac{\theta(x,a,C)}{\theta'(y,b,C)} \rightarrow 0$, as $R \rightarrow \infty$, implies that for any $K_0 > 0$, we can select sufficiently large R such that

$$(28) \quad \ln \frac{\theta(x, a, C)}{\theta'(y, b, C)} \leq -2K_0 - \ln 2, \forall x, y \in B,$$

so that

$$(29) \quad \frac{Q_{x,a}[e^v \mathbf{1}_{C^c}]}{Q_{y,b}[e^v \mathbf{1}_C]} \leq e^{2(K-K_0) + \mathcal{U}_x(w_0) - \mathcal{U}_y(-w_0) - \ln 2}$$

where $\mathcal{U}_x(v) := \sup_{a \in A(x)} \mathcal{R}_{x,a}(v)$ (cf. the proof of Theorem 3.14 in the Appendix). Now we consider the first quotient in (27). By Assumption 4.6, we immediately have $\frac{Q_{x,a}[e^v \mathbf{1}_C]}{Q_{y,b}[e^v \mathbf{1}_C]} \leq \frac{\lambda_C^+}{\lambda_C^-}$. Hence, setting

$$(30) \quad K := K_0 + \frac{1}{2} \ln 2 + \ln \left(\frac{\lambda_C^+}{\lambda_C^-} \right),$$

we obtain $Q_{x,a}[e^v \mathbf{1}_C]/Q_{y,b}[e^v \mathbf{1}_C] \leq e^{2(K-K_0) + \mathcal{U}_x(w_0) - \mathcal{U}_y(-w_0) - \ln 2}$. Together with (29), it yields the required inequality:

$$\frac{Q_{x,a}[e^v]}{Q_{y,b}[e^v]} \leq e^{2(K-K_0) + \mathcal{U}_x(w_0) - \mathcal{U}_y(-w_0)},$$

where K is chosen according to (30), while R is determined by (28). \square

We now investigate the Doeblin-type condition (B2) stated in Section 3.3 for the upper envelope $\bar{\mathcal{R}}^{(w,K)}$ of the entropic map.

PROPOSITION 4.9. *Let $w : \mathsf{X} \rightarrow [1, \infty)$ be a $\mathcal{B}(\mathsf{X})$ -measurable function and $\mathsf{B} := \mathsf{B}_w(R)$ with some $R > 0$. Suppose Assumption 4.6 holds. Assume further that $\bar{\mathcal{R}}_{x,a}^{(w,K)}(w_0) < \infty$ for all $x \in \mathsf{B}$ and $a \in \mathsf{A}(x)$. Then there exist a constant $\alpha \in (0, 1)$ and a probability measure on $(\mathsf{X}, \mathcal{B}(\mathsf{X}))$ satisfying*

$$\bar{\mathcal{R}}_{x,a}^{(w,K)}(v) - \bar{\mathcal{R}}_{x,a}^{(w,K)}(u) \geq \alpha \mu[v - u], \forall x \in \mathsf{B}, a \in \mathsf{A}(x), v \geq u \in \mathcal{B}_{1+w_0}.$$

Proof. Note that since $\bar{\mathcal{R}}_{x,a}^{(w,K)}(w_0) < \infty$, we have for all $v \in \mathcal{B}_{1+w_0}$, $x \in \mathsf{B}$ and $a \in \mathsf{A}(x)$,

$$|\bar{\mathcal{R}}_{x,a}^{(w,K)}(v)| \leq \bar{\mathcal{R}}_{x,a}^{(w,K)}(|v|) \leq \|v\|_{1+w_0} \bar{\mathcal{R}}_{x,a}^{(w,K)}(1+w_0) < \infty.$$

By (24), we have for all $x \in \mathsf{B}$ and $v \geq u \in \mathcal{B}_{1+w_0}$,

$$\begin{aligned} \bar{\mathcal{R}}_{x,a}^{(w,K)}(v) - \bar{\mathcal{R}}_{x,a}^{(w,K)}(u) &= \sup_{h \in \mathcal{B}_w^{(K)}} \frac{Q_{x,a}[e^h v]}{Q_{x,a}[e^h]} - \sup_{h' \in \mathcal{B}_w^{(K)}} \frac{Q_{x,a}[e^{h'} u]}{Q_{x,a}[e^{h'}]} \\ &\geq \inf_{h' \in \mathcal{B}_w^{(K)}} \frac{Q_{x,a}[e^{h'}(v - u)]}{Q_{x,a}[e^{h'}]}. \end{aligned}$$

By Remark 4.7, Assumption 4.6 implies that there exist a probability measure μ_{B} and α_{B} such that $Q_{x,a}[v] \geq \alpha_{\mathsf{B}} \mu_{\mathsf{B}}[v]$ for all nonnegative measurable functions v . Hence, for all $x \in \mathsf{B}$ and $h' \in \mathcal{B}_w^{(K)}$, we have

$$\begin{aligned} \frac{Q_{x,a}[e^{h'}(v - u)]}{Q_{x,a}[e^{h'}]} &\geq \frac{\alpha_{\mathsf{B}} \mu_{\mathsf{B}}[e^{h'}(v - u)]}{Q_{x,a}[e^{Kw}]} \geq \frac{\alpha_{\mathsf{B}} \mu_{\mathsf{B}}[e^{-Kw}(v - u)]}{\max_{x \in \mathsf{B}} Q_{x,a}[e^{Kw}]} \\ &= \frac{\alpha_{\mathsf{B}} \mu_{\mathsf{B}}[e^{-Kw}]}{\max_{x \in \mathsf{B}} Q_{x,a}[e^{Kw}]} \frac{\mu_{\mathsf{B}}[e^{-Kw}(v - u)]}{\mu_{\mathsf{B}}[e^{-Kw}]}. \end{aligned}$$

Hence, $\alpha := \frac{\alpha_{\mathsf{B}} \mu_{\mathsf{B}}[e^{-Kw}]}{\max_{x \in \mathsf{B}} Q_{x,a}[e^{Kw}]}$ and the probability measure $d\mu := \frac{e^{-Kw} d\mu_{\mathsf{B}}}{\int e^{-Kw} d\mu_{\mathsf{B}}}$ are the required constant and probability measure respectively. \square

The following theorem shows that applying the entropic map, along with an additional growth condition for cost functions (see (31) below), the existence of Lyapunov function stated in Assumption 4.3 and the local Doeblin condition stated in Assumption 4.6 are sufficient for Assumption 3.6 and 3.12.

THEOREM 4.10. *Let \mathcal{R} be the entropic map with $\lambda = 1$. Suppose Assumption 4.3 and 4.6 hold with a weight function w_1 . If the cost function c satisfies*

$$(31) \quad |c(x, a)| \leq C w_1^q(x), \forall (x, a) \in \mathsf{K}, \text{ with some } q \in (0, 1),$$

then Assumption 3.6 holds with $w_0 = w_1^p$ for any $p \in (q, 1)$, and some $K > 0$, and Assumption 3.12 holds with $w = 1 + K^{-1}w_0$.

Proof. Fix one $p \in (q, 1)$ and let $w_0 = w_1^p$. Then by Corollary 4.5, there exists $\hat{\gamma}_0 \in (0, 1)$ and $\hat{K}_0 > 0$ satisfying $\mathcal{U}_x(w_0) \leq \hat{\gamma}_0 w_0(x) + \hat{K}_0$. Choosing one $\gamma_0^{(c)} \in (0, 1 - \hat{\gamma}_0)$, there exists a constant $K_0^{(c)} > 0$ satisfying $Cw_1^q(x) \leq \gamma_0^{(c)} w_0(x) + K_0^{(c)}$. Hence, Assumption 3.6 (A1) holds with $\gamma_0 := \hat{\gamma}_0 + \gamma_0^{(c)} \in (0, 1)$ and $K_0 := \hat{K}_0 + K_0^{(c)}$. Due to Remark 4.7, Assumption 3.6 (A2) holds with some constant $K > K_0$. Next, by Theorem 3.7 and 4.4, Assumption 3.12 (B1) holds with $w := 1 + K^{-1}w_0$. Assumption 3.12 (B2) holds due to Proposition 4.9. \square

4.3 Discussion

Most of the existing literature on risk-sensitive MCPs applying the entropic map, considers finite or countable state spaces [3, 5, 6, 10, 18, 23, 35] or bounded cost functions [4]. We consider in this paper a more general setting: Borel state-action spaces and unbounded cost functions. The literature under the same setting includes [13, 29, 30, 31] and [14], with which we provide below a detailed comparison.

Among others, Kontoyiannis and Meyn [31] developed (see also their earlier work [30]) a spectral theory of multiplicative Markov processes, where the Poisson equation w.r.t. the entropic map (called multiplicative Poisson equation in [31]) plays the central role. Though our assumptions are less general than the assumptions stated in [30, 31], our proof that generalizes the Hairer-Mattingly approach [22] is conceptually simpler than the one provided in [30, 31], and can also be applied to other types of valuation maps.

To guarantee the existence of a solution to the optimal control problem, Jaśkiewicz stated a set of more general conditions in [29] than the set stated in this section. However, Condition (B) in [29] which assumes the boundedness of iterations is very difficult to verify. In contrast to it, we provide in this section a set of more verifiable conditions that guarantees this boundedness (see Theorems 4.10 and 3.7) and hence Condition (B) in [29]. It is also worth to mention that the cost functions allowed in [29] are assumed to be lower bounded (but not upper bounded), which is not as general as our setting.

The assumption (A4) in Section 4 of Di Masi and Stettner [14] (see also their earlier work [13]) requires a positive continuous density, i.e., there exists a positive function q satisfying $Q(dy|x, a) = q(x, a, y)\mu(dy)$ for some reference probability measure μ , which implies the local Doeblin condition in Assumption 4.6. Hence, our assumption is more general than its counterpart in [14]. The assumption (A3) set in Section 3 of [14] for the cost function c is implicit and difficult to be verified. On the contrary, the sufficient growth condition (31) for c , is explicit in form of the weight function w_1 w.r.t. the entropic map. Note that, in the example provided by [14], the assumption (A3) is also verified with the help of a weight function.

Finally, as an advantage, in comparison with [29, 31] and [14], the convergence rate of iterations towards the solution to the Poisson equation is explicitly specified by $\bar{\alpha}$ in Theorem 3.14 under the chosen seminorm.

5 Examples

In this section, we utilize the following example taken from [14, Section 6] as a canonical example of MCPs. Let $\|\cdot\|$ denote the Euclidean norm, and $\|\cdot\|_\infty$ the sup-norm.

5.1 A canonical example of MCP

Let $\mathsf{X} = \mathbb{R}^d$ and A be a compact subset of a Euclidean space. Consider the following discretized ergodic diffusion $\{x_t \in \mathbb{R}^d\}$:

$$(32) \quad X_{t+1} = AX_t + b(X_t, A_t) + D(X_t, A_t)W_t, \text{ where}$$

- $\{W_t \in \mathbb{R}^d\}$ is a sequence of i.i.d. standard white noise;
- $D : \mathsf{K} \rightarrow \mathbb{R}^{d \times d}$ is a continuous bounded matrix-valued function which is uniformly elliptic, i.e., there exists a constant $L > 0$ such that

$$(33) \quad L^{-1}\|\xi\|^2 \leq \xi^\top D^\top(x, a)D(x, a)\xi \leq L\|\xi\|^2, \forall (x, a) \in \mathsf{K}, \xi \in \mathbb{R}^d;$$

- $b : \mathsf{K} \rightarrow \mathbb{R}^d$ is a continuous bounded vector function, i.e., there exists a positive constant $B > 0$ such that $\|b(x, a)\|^2 \leq B, \forall (x, a) \in \mathsf{K}$; and
- A is a matrix satisfying that there exists a constant $\tilde{\gamma} \in (0, 1)$ such that $\xi^\top A^\top A \xi \leq \tilde{\gamma}\|\xi\|^2, \forall \xi \in \mathbb{R}^d$.

Then the transition kernel $Q(dy|x, a)$ has the following density w.r.t. the Lebesgue measure,

$$(34) \quad q(y|x, a) = (2\pi)^{-d/2} |\Sigma|^{1/2} e^{-\frac{1}{2}(y-Ax-b)^\top \Sigma (y-Ax-b)}, \text{ with } \Sigma := (DD^\top)^{-1}.$$

Remark 5.1. Note that since for each x and y , $a \mapsto q(y|x, a)$ is continuous on $\mathsf{A}(x)$, $Q_{x,a}$ is weakly continuous on $\mathsf{A}(x)$, i.e., for each $x \in \mathsf{X}$ and bounded $v \in \mathcal{B}$, $a \mapsto Q_{x,a}(v)$ is continuous on $\mathsf{A}(x)$.

5.2 Entropic maps

Let \mathcal{R} be the entropic map defined in Example 2.7 with $\lambda = 1$. By Theorem 4.10, it is sufficient to verify the existence of Lyapunov function stated in Assumption 4.3 and the local Doeblin condition stated in Assumption 4.6.

Among them, the local Doeblin condition is satisfied since the transition kernel Q has a positive continuous density function q w.r.t. the Lebesgue measure.

It remains to verify the the existence of Lyapunov function stated in Assumption 4.3. Take one $\gamma \in (\tilde{\gamma}, 1)$ and consider the following weight function

$$(35) \quad \hat{w}_1(x) = \frac{\epsilon}{2}\|x\|^2, \text{ with some positive } \epsilon \leq \frac{\gamma - \tilde{\gamma}}{\gamma} L^{-1} < L^{-1}.$$

Hence, $\Sigma(x, a) - \epsilon I$ is positive definite for all $(x, a) \in \mathbf{K}$. We show that \hat{w}_1 is a Lyapunov function w.r.t. the entropic map as follows. By setting $\tilde{x} := Ax + b$, we obtain

$$\begin{aligned} \int Q(dy|x, a) e^{\hat{w}_1(y)} &= (2\pi)^{-d/2} |\Sigma|^{1/2} \int e^{-\frac{1}{2}(y^\top (\Sigma - \epsilon I)y - 2y^\top \Sigma \tilde{x}^\top + \tilde{x}^\top \Sigma \tilde{x})} dy \\ &= \frac{|\Sigma|^{1/2}}{|\Sigma - \epsilon I|^{1/2}} e^{\frac{1}{2} \tilde{x}^\top \Sigma ((\Sigma - \epsilon I)^{-1} - \Sigma^{-1}) \Sigma \tilde{x}}, \text{ which yields} \end{aligned}$$

$$(36) \quad Q_{x,a}[e^{\hat{w}_1}] = \frac{|\Sigma|^{1/2}}{|\Sigma - \epsilon I|^{1/2}} e^{\frac{1}{2} (Ax+b)^\top \Sigma ((\Sigma - \epsilon I)^{-1} - \Sigma^{-1}) \Sigma (Ax+b)}.$$

By (33) and the choice of ϵ in (35), we have

$$\frac{1}{2} x^\top A^\top \Sigma ((\Sigma - \epsilon I)^{-1} - \Sigma^{-1}) \Sigma Ax \leq \frac{\gamma \epsilon}{2} \|x\|^2 = \gamma \hat{w}_1(x), \forall (x, a) \in \mathbf{K}.$$

Finally, due to the uniform boundedness of b and $\frac{|\Sigma|^{1/2}}{|\Sigma - \epsilon I|^{1/2}}$, we can always select some $\gamma_1 \in (\gamma, 1)$ and $\hat{K} > 0$ such that

$$\ln \left(\int Q(dy|x, a) e^{\hat{w}_1(y)} \right) \leq \gamma_1 \hat{w}_1(x) + \hat{K}, \forall (x, a) \in \mathbf{K}.$$

Hence, Assumption 4.3 is verified with $w_1 := 1 + \hat{w}_1$, γ_1 and $K_1 := \hat{K}_1 + 1 - \gamma$.

A final remark is addressed to Assumption (C3), where the lower semicontinuity can be in fact strengthen to be continuity in this example. Note that for each $x \in \mathbf{X}$, (i) $v \in \mathcal{B}$, $a \mapsto Q_{x,a}[v]$ is continuous on $\mathbf{A}(x)$ (see Remark 5.1); and (ii) by (36), $a \mapsto Q_{x,a}[\hat{W}]$ is continuous on $\mathbf{A}(x)$, where $\hat{W} := e^{\hat{w}_1}$. Hence, by [26, Lemma 8.3.7(a)], for each $v \in \mathcal{B}_{\hat{W}}$, $a \mapsto Q_{x,a}[v]$ is continuous on $\mathbf{A}(x)$. Note that if $v \in \mathcal{B}_{w_0}$ satisfies $|v| \leq w_0 + K$, where $w_0 = w_1^p$, $p \in (0, 1)$ (see Theorem 4.10), then $e^v \in \mathcal{B}_{\hat{W}}$. Hence, we obtain immediately the required weak continuity.

5.3 Mean-semideviation trade-off

Recall that in Example 2.9 the mean-semideviation tradeoff is defined as

$$\mathcal{R}_{x,a}(v) := Q_{x,a}[v] + \lambda \left(Q_{x,a} \left[(v - Q_{x,a}[v])_+^r \right] \right)^{1/r}, \lambda \in (0, 1).$$

Note that since \mathcal{R} is coherent [43], Assumptions 3.6 and 3.12 can be reduced to Assumption 3.9 (see Remark 3.13). We consider below the case $r = 2$.

Verification of (A1a') Consider the canonical example. Let $\tilde{x} := Ax + b(x, a)$ and set $w_0(x) = \|x\|^2$. Hence, given (x, a) , $Y := \tilde{x} + D(x, a)W$, where W is a d -dimensional white noise, follows a multivariate normal distribution with mean \tilde{x} and

covariance matrix DD^\top . We have then

$$\begin{aligned} Q_{x,a}[w_0] &= Q_{x,a}[\|Y\|^2] = \|\tilde{x}\|^2 + \text{tr}(DD^\top), \text{ and} \\ Q_{x,a}[w_0 - Q_{x,a}[w_0]]^2 &= Q_{x,a}[w_0^2] - (Q_{x,a}[w_0])^2 \\ &= 4\mathbb{E}[\tilde{x}^\top DW]^2 + 4\mathbb{E}[\tilde{x}^\top DW\|DW\|^2] \\ &\quad + \mathbb{E}[\|DW\|^4] - (\mathbb{E}[\|DW\|^2])^2. \end{aligned}$$

Here, the expectation \mathbb{E} is taken over the white noise W . By the assumptions on A and b , there exist constants $\epsilon \in (\tilde{\gamma}, 1)$, $L_0 > 0$ such that $\|\tilde{x}\|^2 \leq \epsilon\|x\|^2 + L_0$. By the assumption on D , there exists a constant $L_1 > 0$ such that

$$Q_{x,a}[w_0 - Q_{x,a}[w_0]]^2 \leq L_1(\|\tilde{x}\|^2 + 1).$$

Hence, we have

$$\begin{aligned} \mathcal{R}_{x,a}(w_0) &= Q_{x,a}[w_0] + \lambda \sqrt{Q_{x,a}[(w_0 - Q_{x,a}[w_0])_+^2]} \\ &\leq Q_{x,a}[w_0] + \lambda \sqrt{Q_{x,a}[(w_0 - Q_{x,a}[w_0])^2]} \\ &\leq \epsilon\|x\|^2 + L_0 + \text{tr}(DD^\top) + \lambda \sqrt{L_1(\|\tilde{x}\|^2 + 1)} \end{aligned}$$

which implies that there exist positive constants L'_0 and L'_1 such that

$$(37) \quad \mathcal{R}_{x,a}(w_0) = \epsilon w_0(x) + L'_0 + \lambda L'_1 \sqrt{w_0(x) + 1}$$

Taking one $\gamma_0 \in (\epsilon, 1)$, since $t \mapsto (\epsilon - \gamma_0)t + L'_0 + \lambda L'_1 \sqrt{1 + t}$ is concave in $[0, \infty)$ and its maximum is attained at some constant K_0 , we have $\mathcal{R}_{x,a}(w_0) \leq \gamma_0 w_0(x) + K_0$, which verifies (A1a').

Verification of (A2') One easy extension of the subgradient calculation presented by Svindland [48] (see also [45, Section 6]) yields

$$\mathcal{R}_{x,a}(v) - \mathcal{R}_{x,a}(u) \geq \int g(x, a, u, y)(v(y) - u(y))Q_{x,a}(\mathrm{d}y), \forall v \geq u \in \mathcal{B}_{1+w_0}, \text{ with}$$

$$g(x, a, u) := \begin{cases} 1 & \text{if } u \text{ is constant} \\ 1 - \lambda \frac{Q_{x,a}[(u - Q_{x,a}[u])_+] - (u - Q_{x,a}u)_+}{\sqrt{Q_{x,a}[(u - Q_{x,a}[u])_+^2]}} & \text{otherwise} \end{cases}$$

On the other hand,

$$Q_{x,a}[(u - Q_{x,a}u)_+] - (u - Q_{x,a}[u])_+ \leq Q_{x,a}[(u - Q_{x,a}u)_+] \leq \sqrt{Q_{x,a}[(u - Q_{x,a}u)_+^2]}$$

implies that $g(x, a, u) \geq 1 - \lambda > 0$. Note that in the canonical MCP, $Q_{x,a}(\cdot)$ is supported by a probability measure μ on any bounded level-set. Hence,

$$\mathcal{R}_{x,a}(v) - \mathcal{R}_{x,a}(u) \geq (1 - \lambda) \int (v(y) - u(y))Q_{x,a}(\mathrm{d}y) \geq \alpha(1 - \lambda)\mu[v - u]$$

verifies (A2').

Verification of (C3) The lower semicontinuity can be in fact strengthen to be continuity in this example. Note that for each $x \in \mathbf{X}$, (i) $Q_{x,a}$ is weakly continuous on $\mathbf{A}(x)$ (see Remark 5.1) and (ii) by $a \mapsto Q_{x,a}[w_0]$ is continuous on $\mathbf{A}(x)$. By [26, Lemma 8.3.7(a)], for each $v \in \mathcal{B}_{1+w_0}$, $a \mapsto Q_{x,a}[v]$ is continuous on $\mathbf{A}(x)$. It remains to show that for each $x \in \mathbf{X}$ and $v \in \mathcal{B}_{1+w_0}$, $a \mapsto \sqrt{Q_{x,a}[(v - Q_{x,a}[v])_+^2]}$ is continuous on $\mathbf{A}(x)$. Since $t \mapsto \sqrt{t}$ is continuous on $[0, \infty)$, it is sufficient to show the continuity of the mapping $a \mapsto Q_{x,a}[(v - Q_{x,a}[v])_+^2]$. Now fix $x \in \mathbf{X}$ and $v \in \mathcal{B}_{1+w_0}$, and let $\{a_i \in \mathbf{A}(x)\}$ converging to $a \in \mathbf{A}(x)$ as $i \rightarrow \infty$. Setting $\mu_i = Q_{x,a_i}$ and $\mu = Q_{x,a}$, we have for each i

$$(38) \quad \begin{aligned} & |\mu[(v - \mu[v])_+^2] - \mu_i[(v - \mu_i[v])_+^2]| \\ & \leq |\mu[(v - \mu[v])_+^2] - \mu_i[(v - \mu[v])_+^2]| + |\mu_i[(v - \mu[v])_+^2] - \mu_i[(v - \mu_i[v])_+^2]| \end{aligned}$$

It is easy to verify that for each $x \in \mathbf{X}$, $a \mapsto Q_{x,a}[w_0^2]$ is also continuous on $\mathbf{A}(x)$. Hence, by [26, Lemma 8.3.7(a)], for each $v \in \mathcal{B}_{1+w_0^2}$, $x \in \mathbf{X}$, $a \mapsto Q_{x,a}[v]$ is continuous on $\mathbf{A}(x)$. This implies that the first term in (38) converges to 0 as $i \rightarrow \infty$, since for each $v \in \mathcal{B}_{1+w_0}$, $(v - \mu[v])_+^2 \in \mathcal{B}_{1+w_0^2}$.

Now we consider the second term in (38). Note that for any $x, y \in \mathbb{R}$, we have $|(x)_+ - (y)_+| \leq |x - y|$. Hence, for any $x \in \mathbf{X}$,

$$|(v(x) - \mu[v])_+^2 - (v(x) - \mu_i[v])_+^2| \leq |\mu[v] - \mu_i[v]| |(v(x) - \mu[v])_+ + (v(x) - \mu_i[v])_+|.$$

It yields

$$\begin{aligned} & |\mu_i[(v - \mu[v])_+^2] - \mu_i[(v - \mu_i[v])_+^2]| \\ & \leq \| (v - \mu[v])_+^2 - (v - \mu_i[v])_+^2 \|_{1+w_0} \mu_i[1 + w_0] \\ & \leq |\mu[v] - \mu_i[v]| \| (v - \mu[v])_+ + (v - \mu_i[v])_+ \|_{1+w_0} (\epsilon w_0(x) + L'_0 + 1) \\ & \leq |\mu[v] - \mu_i[v]| (2\|v\|_{1+w_0} + 2(\epsilon w_0(x) + L'_0 + 1)) (\epsilon w_0(x) + L'_0 + 1), \end{aligned}$$

where in the last two inequalities we use the fact that $\mu_i[w_0] \leq \epsilon w_0(x) + L'_0$ (cf. (37)), $\forall i \in \mathbb{N}$. Finally, by the fact that $\forall v \in \mathcal{B}_{1+w_0}$, $\mu[v] \rightarrow \mu_i[v]$ as $i \rightarrow \infty$, we obtain the convergence of the second term in (38).

5.4 Utility-based shortfall

Consider the utility-based shortfall defined in (12)

$$(39) \quad \mathcal{R}_{x,a}(v) = \sup \left\{ m \in \mathbb{R} \mid \int_{\mathbf{X}} u(v(y) - m) Q_{x,a}(dy) \geq 0 \right\},$$

under the assumption that u is increasing and there exist constants l and L satisfying $0 < l \leq 1 \leq L < \infty$ and

$$(40) \quad l \leq \frac{u(x) - u(y)}{x - y} \leq L, \forall x, y \in \mathbb{R}.$$

In other words, for each $x, y \in \mathbb{R}$, we have $u(x) - u(y) = \delta(x, y)(x - y)$, with some $\delta(x, y) \in [l, L]$.

Remark 5.2. Note that u is not required to be convex, nor concave. Hence, the induced risk preference can be mixed and is, therefore, useful for quantifying human behavior [46]. One example of u that satisfies the assumption we made above is a piecewise linear function with slopes upper bounded by L and lower bounded by l .

Föllmer and Schied [20, Proposition 4.104] show that if u is continuous and strictly increasing, the optimal m^* is obtained when the equality holds, i.e., for each $(x, a) \in \mathbb{K}$, $m^*(x, a) := \mathcal{R}_{x,a}(v)$ satisfies

$$(41) \quad \int u(v(y) - m^*(x, a)) Q_{x,a}(dy) = 0.$$

Let $m = \mathcal{R}(v)$ and $m' = \mathcal{R}(v')$. Hence, for each $(x, a) \in \mathbb{K}$,

$$\int u(v(y) - m(x, a)) Q_{x,a}(dy) = \int u(v'(y) - m'(x, a)) Q_{x,a}(dy) = 0.$$

Let $k := (x, a)$. Then,

$$\begin{aligned} 0 &= \int u(v(y) - m(k)) Q_k(dy) - \int u(v'(y) - m'(k)) Q_k(dy) \\ &= \int \delta(v, v', k, y) (v(y) - v'(y) - m(k) + m'(k)) Q_k(dy), \end{aligned}$$

where $\delta(v, v', k, y) := \frac{u(v(y)-m(k))-u(v'(y)-m'(k))}{v(y)-v'(y)-m(k)+m'(k)} \in [l, L]$. Hence,

$$(42) \quad (m(k) - m'(k)) \int \delta(v, v', k, y) Q_k(dy) = \int \delta(v, v', k, y) (v(y) - v'(y)) Q_k(dy).$$

Verification of Assumption 3.6 Let $w_0(x) = e^{\epsilon \|x\|^2}$ be a weight function and we have shown in Section 5.2 that it satisfies $Q_{x,a}[w_0] \leq C(w_0(x))^\gamma$, for some $C > 0$ and $\gamma \in (0, 1)$. First, taking $v = w_0$ and $v' = 0$ in (42), we have $m' = 0$ and $\mathcal{R}_{x,a}(w_0) = m(x) \leq \frac{L}{l} Q_{x,a}[w_0] \leq \frac{L}{l} C(w_0(x))^\gamma$. Second, taking $v = 0$ and $v' = -w_0$ in (42), we have $m = 0$ and

$$(43) \quad -\mathcal{R}_{x,a}(-w_0) = -m'(x, a) \leq \frac{L}{l} Q_{x,a}[w_0] \leq \frac{L}{l} C(w_0(x))^\gamma.$$

Hence, given a cost function c satisfying $|c(x, a)| \leq C'(1 + (w_0(x))^{\gamma'})$, $\forall (x, a) \in \mathbb{K}$, with some $\gamma' \in (0, 1)$ and $C' > 0$, we can always find a sufficiently large K_0 such that

$$(44) \quad (c(x, a) + \mathcal{R}_{x,a}(w_0)) \vee (-c(x) - \mathcal{R}_{x,a}(-w_0)) \leq \gamma_0 w_0(x) + K_0, \forall (x, a) \in \mathbb{K}.$$

This verifies Assumption 3.6 (A1).

Take $v' = w_0 + K$ in (42), where K will be specified later. Due to the assumption

$v \leq w_0 + K$, we have $m \leq m'$ and for any $k = (x, a) \in \mathbf{K}$,

$$\begin{aligned}
L(m(k) - m'(k)) &\leq (m(k) - m'(k)) \int \delta(v, v', k, y) Q_k(dy) \\
&\leq \int \delta(v, v', k, y) (v(y) - v'(y)) Q_k(dy) \\
&\leq l \int (v(y) - v'(y)) Q_k(dy) = l \int (v(y) - w_0(y) - K) Q_k(dy)
\end{aligned}$$

which implies $\mathcal{R}_k(v) - \mathcal{R}_k(w_0 + K) \leq \frac{l}{L} \int (v(y) - w_0(y) - K) Q_k(dy)$.

Analogously we obtain $\mathcal{R}_k(-w_0 - K) - \mathcal{R}_k(v) \leq \frac{l}{L} \int (-w_0 - K - v(y)) Q_k(dy)$. On the other hand, for the canonical MCP, we have $Q_{x,a}(\cdot) \geq \alpha \mu(\cdot)$ with some probability measure μ and $\alpha > 0$ for each x in any bounded level-set and $a \in \mathbf{A}(x)$. Hence, $\frac{l}{L} Q_{x,a}[w_0 + K - v] + \frac{l}{L} Q_{x',a'}[w_0(y) + K + v] \geq \frac{l}{L} 2\alpha K$ yields

$$\mathcal{R}_{x,a}(w + K) - \mathcal{R}_{x,a}(v) + \mathcal{R}_{x',a'}(v) - \mathcal{R}_{x',a'}(-w_0 - K) \geq 2\frac{\alpha l}{L} K.$$

Therefore, taking $K := \frac{L}{\alpha l} K_0$, the Assumption 3.6 (A2) holds.

Verification of Assumption 3.12 By (42), we have

$$\begin{aligned}
\mathcal{R}_{x,a}(v) - \mathcal{R}_{x,a}(v') &\leq \frac{\int \delta(v, v', x, a, y) Q_{x,a}(dy) (v(y) - v'(y))}{\int \delta(v, v', x, a, y) Q_{x,a}(dy)} \\
(45) \quad &\leq \sup_{\delta: l \leq \delta(x, a, y) \leq L} \frac{\int \delta(x, a, y) Q_{x,a}(dy) (v(y) - v'(y))}{\int \delta(x, a, y) Q_{x,a}(dy)} =: \bar{\mathcal{R}}_{x,a}(v - v')
\end{aligned}$$

It is easy to see that $\bar{\mathcal{R}}$ is a coherent risk map. Note that

$$\bar{\mathcal{R}}_{x,a}(w_0) = \sup_{\delta: l \leq \delta(x, a, y) \leq L} \frac{\int \delta(x, a, y) w(y) Q_{x,a}(dy)}{\int \delta(x, a, y) Q_{x,a}(dy)} \leq \frac{L}{l} Q_{x,a}[w_0]$$

which implies that w_0 satisfies (B1) with some constants $\gamma \in (0, 1)$ and $\bar{K} > 0$.

On the other hand, for any $v \geq v' \in \mathcal{B}_{1+w_0}$,

$$\begin{aligned}
\bar{\mathcal{R}}_{x,a}(v) - \bar{\mathcal{R}}_{x,a}(v') &\geq \inf_{\delta: l \leq \delta(x, a, y) \leq L} \frac{\int \delta(x, a, y) Q_{x,a}(dy) (v(y) - v'(y))}{\int \delta(x, a, y) Q_{x,a}(dy)} \\
(46) \quad &\geq \frac{l}{L} Q_{x,a}[v - v'] \geq \frac{\alpha l}{L} \mu[v - v'],
\end{aligned}$$

holds for all x in any bounded level-set and $a \in \mathbf{A}(x)$. Hence, (B2) holds.

Verification of (C3) The lower semicontinuity can be in fact strengthened to be continuity in this example. Let $v \in \mathcal{B}_{1+w_0}$ satisfying $|v| \leq w_0 + K$. Fix $x \in \mathbf{X}$, and let $\{a_i\}$ be a sequence of actions in $\mathbf{A}(x)$ converging to a . Set $\mu_i = Q_{x,a_i}$, $\mu = Q_{x,a}$,

$m_i = \mathcal{R}_{x,a_i}(v)$ and $m = \mathcal{R}_{x,a}(v)$. It is therefore to show m_i converges to m . Indeed, by (41), we have

$$\begin{aligned}
0 &= \int u(v(y) - m_i) \mu_i(dy) - \int u(v(y) - m) \mu(dy) \\
&= \int (u(v(y) - m_i) - u(v(y) - m)) \mu_i(dy) + \int u(v(y) - m) (\mu_i(dy) - \mu(dy)) \\
&= \int \delta_i(y) (m - m_i) \mu_i(dy) + \int u(v(y) - m) (\mu_i(dy) - \mu(dy)) \\
&= (m - m_i) \int \delta_i(y) \mu_i(dy) + \int u(v(y) - m) (\mu_i(dy) - \mu(dy))
\end{aligned}$$

where $\delta_i(y) := \frac{u(v(y)-m_i)-u(v(y)-m)}{m_i-m} \in [l, L]$. Hence,

$$|m - m_i| = \frac{|\int u(v(y) - m) (\mu_i(dy) - \mu(dy))|}{\int \delta_i(y) \mu_i(dy)} \leq \frac{1}{l} \left| \int u(v(y) - m) (\mu_i(dy) - \mu(dy)) \right|.$$

Note that by $|u(v(y) - m)| = |u(v(y) - m) - u(0)| \leq L|v(y) - m|$, we have

$$\|u(v(\cdot) - m)\|_{1+K^{-1}w_0} \leq L(\|v\|_{1+K^{-1}w_0} + m)$$

and hence $u(v(\cdot) - m) \in \mathcal{B}_{1+K^{-1}w_0}$. On the other hand, since for each $x \in \mathbf{X}$, (i) $a \mapsto Q_{x,a}[v]$ is continuous for each $v \in \mathcal{B}$ (see Remark 5.1) and (ii) $a \mapsto Q_{x,a}[w_0]$ is continuous, by Lemma 8.3.7(a) in [26], we have for any $v \in \mathcal{B}_{1+K^{-1}w_0}$, $\mu_i[v] \rightarrow \mu[v]$ as $i \rightarrow \infty$. This implies the convergence of m_i to m .

References

- [1] A. Arapostathis, V.S. Borkar, E. Fernández-Gaucherand, M.K. Ghosh, and S.I. Marcus. Discrete-time controlled Markov processes with average cost criterion: a survey. *SIAM Journal on Control and Optimization*, 31(2):282–344, 1993.
- [2] P. Artzner, F. Delbaen, J.M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.
- [3] G. Avila-Godoy and E. Fernández-Gaucherand. Controlled Markov chains with exponential risk-sensitive criteria: modularity, structured policies and applications. In *Proceedings of the 37th IEEE Conference on Decision and Control*, pages 778–783, 1998.
- [4] N. Bäuerle and U. Rieder. More risk-sensitive Markov decision processes. *Mathematics of Operations Research*, 39, 2013.
- [5] V. S. Borkar and S.P. Meyn. Risk-sensitive optimal control for Markov decision processes with monotone cost. *Mathematics of Operations Research*, pages 192–209, 2002.
- [6] R. Cavazos-Cadena. Optimality equations and inequalities in a class of risk-sensitive average cost Markov decision chains. *Mathematical Methods of Operations Research*, 71(1):47–84, 2010.

- [7] Ö. Çavuş and A. Ruszczyński. Risk-averse control of undiscounted transient Markov models. *SIAM Journal on Control and Optimization*, 52(6):3935–3966, 2013.
- [8] P. Cheridito and T. Li. Risk measures on Orlicz hearts. *Mathematical Finance*, 19(2): 189–214, 2009.
- [9] K.J. Chung and M.J. Sobel. Discounted MDPs: distribution functions and exponential utility maximization. *SIAM Journal on Control and Optimization*, 25:49, 1987.
- [10] S.P. Coraluppi and S.I. Marcus. Mixed risk-neutral/minimax control of discrete-time, finite-state Markov decision processes. *IEEE Transactions on Automatic Control*, 45 (3):528–532, 2000.
- [11] P. Del Moral, M. Ledoux, and L. Miclo. On contraction properties of Markov kernels. *Probability Theory and Related Fields*, 126(3):395–420, 2003.
- [12] F. Delbaen. Coherent risk measures on general probability spaces. *Advances in Finance and Stochastics Essays in Honour of Dieter Sondermann*, pages 1–37, 2000.
- [13] G.B. Di Masi and L. Stettner. Infinite horizon risk sensitive control of discrete time Markov processes with small risk. *Systems & control letters*, 40(1):15–20, 2000.
- [14] G.B. Di Masi and L. Stettner. Infinite horizon risk sensitive control of discrete time Markov processes under minorization property. *SIAM Journal on Control and Optimization*, 46(1):231, 2008.
- [15] R. Douc, E. Moulines, and J.S. Rosenthal. Quantitative bounds on convergence of time-inhomogeneous Markov chains. *Annals of Applied Probability*, pages 1643–1665, 2004.
- [16] R. Douc, G. Fort, E. Moulines, and P. Priouret. Forgetting the initial distribution for hidden Markov models. *Stochastic Processes and Their Applications*, 119(4):1235–1256, 2009.
- [17] J.A. Filar, LCM Kallenberg, and H.M. Lee. Variance-penalized Markov decision processes. *Mathematics of Operations Research*, pages 147–161, 1989.
- [18] W.H. Fleming and D. Hernández-Hernández. Risk-sensitive control of finite state machines on an infinite horizon I. *SIAM Journal on Control and Optimization*, 35(5): 1790–1810, 1997.
- [19] H. Föllmer and A. Schied. Convex measures of risk and trading constraints. *Finance and Stochastics*, 6(4):429–447, 2002.
- [20] H. Föllmer and A. Schied. *Stochastic Finance*. Walter de Gruyter & Co., Berlin, 2004. Extended edition.
- [21] S. Gaubert and J. Gunawardena. The Perron-Frobenius theorem for homogeneous, monotone functions. *Transactions American Mathematical Society*, 356(12):4931–4950, 2004.
- [22] M. Hairer and J.C. Mattingly. Yet another look at Harris’ ergodic theorem for Markov chains. In *Seminar on Stochastic Analysis, Random Fields and Applications VI*, pages 109–117. Springer, 2011.

- [23] D. Hernández-Hernández and S.I. Marcus. Risk sensitive control of Markov processes in countable state space. *Systems & Control Letters*, 29(3):147–155, 1996.
- [24] O. Hernández-Lerma. *Adaptive Markov Control Processes*. Springer, 1989.
- [25] O. Hernández-Lerma and J.B. Lasserre. *Discrete-time Markov Control Processes: Basic Optimality Criteria*. Springer, 1996.
- [26] O. Hernández-Lerma and J.B. Lasserre. *Further Topics on Discrete-Time Markov Control Processes*. Springer Verlag, 1999.
- [27] R.A. Howard and J.E. Matheson. Risk-sensitive Markov decision processes. *Management Science*, 18(7):356–369, 1972.
- [28] G.N. Iyengar. Robust dynamic programming. *Mathematics of Operations Research*, pages 257–280, 2005.
- [29] A. Jaśkiewicz. Average optimality for risk-sensitive control with general state space. *The Annals of Applied Probability*, 17(2):654–675, 04 2007.
- [30] I. Kontoyiannis and S.P. Meyn. Spectral theory and limit theorems for geometrically ergodic Markov processes. *Annals of Applied Probability*, pages 304–362, 2003.
- [31] I. Kontoyiannis and S.P. Meyn. Large deviations asymptotics and the spectral theory of multiplicatively regular Markov processes. *Electron. J. Probab.*, 10(3):61–123, 2005.
- [32] M. Kupper and W. Schachermayer. Representation results for law invariant time consistent functions. *Mathematics and Financial Economics*, 2, 209.
- [33] M. Ledoux. *The Concentration of Measure Phenomenon*. American Mathematical Society, 2001.
- [34] D.A. Levin, Y. Peres, and E.L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2009.
- [35] S.I. Marcus, E. Fernández-Gaucherand, D. Hernández-Hernandez, S. Coraluppi, and P. Fard. Risk sensitive Markov decision processes. *Progress in Systems and Control Theory*, 22:263–280, 1997.
- [36] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag London Ltd., London, 1993.
- [37] A. Nilim and L. El Ghaoui. Robust control of markov decision processes with uncertain transition matrices. *Operations Research*, 53(5):780–798, 2005.
- [38] W. Ogryczak and A. Ruszczyński. From stochastic dominance to mean-risk models: Semideviations as risk measures. *European Journal of Operational Research*, 116(1): 33–50, 1999.
- [39] M.L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1994.
- [40] R.T. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

- [41] B. Roorda and J.M. Schumacher. Time consistency conditions for acceptability measures, with an application to tail value at risk. *Insurance: Mathematics and Economics*, 40(2):209–230, 2007.
- [42] A. Ruszczyński. Risk-averse dynamic programming for Markov decision processes. *Mathematical Programming*, pages 1–27, 2010.
- [43] A. Ruszczyński and A. Shapiro. Optimization of convex risk functions. *Mathematics of operations research*, 31(3):433–452, 2006.
- [44] A. Schied, H. Föllmer, and S. Weber. Robust preferences and robust portfolio choice. *Handbook of Numerical Analysis*, 15:29–87, 2009.
- [45] Y. Shen, W. Stannat, and K. Obermayer. Risk-sensitive Markov Control Processes. *SIAM Journal on Control and Optimization*, 51(5):3652–3672, 2013.
- [46] Y. Shen, M.J. Tobia, T. Sommer, and K. Obermayer. Risk-sensitive reinforcement learning. *Neural Computation*, 26(7):1298–1328, July 2014.
- [47] M.J. Sobel. The variance of discounted Markov decision processes. *Journal of Applied Probability*, pages 794–802, 1982.
- [48] G. Svindland. Subgradients of law-invariant convex risk measures on L^1 . *Statistics & Decisions*, 27(2):169–199, 2009.
- [49] A. Tversky and D. Kahneman. Advances in prospect theory: cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

Appendix

Proof of Theorem 3.14 Define $w' := 1 + \beta w_0$ for some positive $\beta \in \mathbb{R}_+$, whose value will be specified later. Suppose $\|v - u\|_{s, w'} = A \geq 0$. Due to Lemma 2.1 and the fact that adding any constant to v and u will not change the values of both sides of the required inequality, we may assume that $\|v - u\|_{w'} = A$. Define

$$\bar{U}_x(v) := \sup_{a \in A(x)} \bar{\mathcal{R}}_{x,a}^{(w,K)}(v), x \in \mathbf{X}.$$

It is easy to verify that $\bar{U}_x(\cdot)$ is monotone, translation invariant, centralized and coherent on $\mathcal{B}_w^{(K)}$ for each $x \in \mathbf{X}$. Hence, we have

$$\mathcal{R}_x^\pi(v) - \mathcal{R}_x^\pi(u) \leq \int \pi(da|x) \bar{\mathcal{R}}_{x,a}^{(w,K)}(v - u) \leq \bar{U}_x(v - u) \leq \bar{U}_x(|v - u|), \forall x \in \mathbf{X}.$$

Switching v and u , we obtain

$$(47) \quad |\mathcal{R}_x^\pi(v) - \mathcal{R}_x^\pi(u)| \leq \bar{U}_x(|v - u|) \leq \|v - u\|_{w'} \bar{U}_x(w'), \forall x \in \mathbf{X}.$$

We consider the following two cases.

Case I: $w_0(x) + w_0(y) \geq R$ and set $\gamma_0 := \gamma + \frac{2\bar{K}}{R} \in (0, 1)$ and $\gamma_1 := \frac{2+\beta R\gamma_0}{2+\beta R}$ for some $\beta > 0$. It is easy to verify that $\gamma_1 \in (0, 1)$. Then (47) yields

$$(48) \quad \begin{aligned} |\mathcal{R}_x^\pi(v) - \mathcal{R}_x^\pi(u) - \mathcal{R}_y^\pi(v) + \mathcal{R}_y^\pi(u)| &\leq |\mathcal{R}_x^\pi(v) - \mathcal{R}_x^\pi(u)| + |\mathcal{R}_y^\pi(v) - \mathcal{R}_y^\pi(u)| \\ &\leq A(2 + \beta \bar{U}_x(w_0) + \beta \bar{U}_y(w_0)) \leq A(2 + \beta \gamma w_0(x) + \beta \gamma w_0(y) + 2\beta \bar{K}) \\ &\leq A(2 + \beta \gamma_0 w_0(x) + \gamma_0 w_0(y)) \leq A\gamma_1(w'(x) + w'(y)). \end{aligned}$$

Case II: $w_0(x) + w_0(y) \leq R$. Hence both x and y are in the subset \mathbf{B} . We define for all $x \in \mathbf{B}$, $\tilde{\mathcal{R}}_x^\pi(v) := \frac{1}{1-\alpha}\mathcal{R}_x^\pi(v) - \frac{\alpha}{1-\alpha}\mu(v)$, and $\tilde{\mathcal{U}}_x(v) := \frac{1}{1-\alpha}\tilde{\mathcal{U}}_x(v) - \frac{\alpha}{1-\alpha}\mu(v)$. Hence, we have $\tilde{\mathcal{R}}_x^\pi(v) - \tilde{\mathcal{R}}_x^\pi(u) \leq \tilde{\mathcal{U}}_x(v-u)$. By Assumption (B2), $\tilde{\mathcal{U}}_x(v) - \tilde{\mathcal{U}}_x(u) \geq 0, \forall v \geq u \in \mathcal{B}_{1+w_0}$, shows that $\tilde{\mathcal{U}}_x(\cdot)$ is monotone. It is also easy to verify that $\tilde{\mathcal{U}}_x(\cdot)$ is translation invariant, centralized and coherent. Hence,

$$\begin{aligned} |\mathcal{R}_x^\pi(v) - \mathcal{R}_x^\pi(u) - \mathcal{R}_y^\pi(v) + \mathcal{R}_y^\pi(u)| &= (1-\alpha)|\tilde{\mathcal{R}}_x^\pi(v) - \tilde{\mathcal{R}}_x^\pi(u) - \tilde{\mathcal{R}}_y^\pi(v) + \tilde{\mathcal{R}}_y^\pi(u)| \\ &\leq (1-\alpha)|\tilde{\mathcal{R}}_x^\pi(v) - \tilde{\mathcal{R}}_x^\pi(u)| + (1-\alpha)|\tilde{\mathcal{R}}_y^\pi(v) - \tilde{\mathcal{R}}_y^\pi(u)| \\ &\leq (1-\alpha)\tilde{\mathcal{U}}_x(|v-u|) + (1-\alpha)\tilde{\mathcal{U}}_y(|v-u|) \\ &\leq 2A(1-\alpha) + A(1-\alpha)\beta \left(\tilde{\mathcal{U}}_x(w_0) + \tilde{\mathcal{U}}_y(w_0) \right). \end{aligned}$$

Note that since $(1-\alpha)\tilde{\mathcal{U}}_x(w_0) \leq \bar{\mathcal{U}}_x(w_0)$ holds for all $x \in \mathbf{B}$, we obtain

$$\begin{aligned} (49) \quad |\mathcal{R}_x^\pi(v) - \mathcal{R}_x^\pi(u) - \mathcal{R}_y^\pi(v) + \mathcal{R}_y^\pi(u)| &\leq 2A(1-\alpha) + A\beta (\bar{\mathcal{U}}_x(w_0) + \bar{\mathcal{U}}_y(w_0)) \\ &\leq 2A(1-\alpha) + A\beta(\gamma(w_0(x) + w_0(y)) + 2\bar{K}). \end{aligned}$$

We select $\beta := \frac{\alpha_0}{K}$ for some $\alpha_0 \in (0, \alpha)$. Setting $\gamma_2 := (1-\alpha + \alpha_0) \vee \gamma \in (0, 1)$ yields for all $x \neq y$

$$\begin{aligned} (50) \quad &|\mathcal{R}_x^\pi(v) - \mathcal{R}_x^\pi(u) - \mathcal{R}_y^\pi(v) + \mathcal{R}_y^\pi(u)| \\ &\leq 2A(1-\alpha + \alpha_0) + A\gamma\beta(w_0(x) + w_0(y)) \leq A\gamma_2(w'(x) + w'(y)). \end{aligned}$$

Hence, setting $\bar{\alpha} := \gamma_1 \vee \gamma_2 < 1$, (48) and (50) imply for all $x \neq y$

$$|\mathcal{R}_x^\pi(v) - \mathcal{R}_x^\pi(u) - \mathcal{R}_y^\pi(v) + \mathcal{R}_y^\pi(u)| \leq \|v-u\|_{s,w'} \bar{\alpha}(w'(x) + w'(y)),$$

the required inequality.

Recall that for any unbounded nonnegative $\mathcal{B}(\mathbf{X})$ -measurable function w and any real number $R \in \mathbb{R}$, we define $\mathbf{B}_w(R) := \{x \in \mathbf{X} | w(x) \leq R\}$ and $\mathbf{B}_w^c(R)$ its complementary set.

Proof of Theorem 4.4 Due to Assumption 4.3, for any $\lambda \in (\gamma_1, 1)$, we have

$$\mathcal{R}_{x,a}(w_1) \leq \lambda w_1(x), \forall x \in \mathbf{B}_{w_1}^c(A), A := \frac{K_1}{\lambda - \gamma_1}, a \in \mathbf{A}(x).$$

It implies that for all $x \in \mathbf{B}_{w_1}^c(A), a \in \mathbf{A}(x)$,

$$\begin{aligned} (51) \quad &\int_{\mathbf{B}_{w_1}^c(\lambda w_1(x))} Q_{x,a}(dy) \left(e^{w_1(y) - \lambda w_1(x)} - 1 \right) \\ &\leq \int_{\mathbf{B}_{w_1}(\lambda w_1(x))} Q_{x,a}(dy) \left(1 - e^{w_1(y) - \lambda w_1(x)} \right). \end{aligned}$$

Taking some $\gamma_2 \in (\lambda^p, 1)$, by the definition of w_0 ($:= w_1^p$), we have then

$$(52) \quad \mathbf{B}_{w_0}^c(\gamma_2 w_0(x)) \subset \mathbf{B}_{w_1}^c(\lambda w_1(x)), \forall x \in \mathbf{X}.$$

Indeed, for any $y \in \mathbf{B}_{w_0}^c(\gamma_2 w_0(x))$, it follows that $w_0(y) > \gamma_2 w_0(x)$, which is equivalent to $w_1(y) > (\gamma_2)^{1/p} w_1(x) > \lambda w_1(x)$. Hence, $y \in \mathbf{B}_{w_1}^c(\lambda w_1(x))$ as well. Before continuing the proof, we state and prove two lemmas.

LEMMA A.3. For any $\eta \in (0, 1 - \lambda)$, $p \in (0, 1)$, $K \geq 0$ and $\gamma_2 \in ((\lambda + \eta)^p, 1)$, there exists a constant $R_1 > 0$ such that for all $y \in \mathcal{B}_{w_0}^c(\gamma_2 w_0(x))$, $x \in \mathcal{B}_{w_1}^c(R)$ and $R \geq R_1$,

$$e^{w_0(y) + K + \eta w_1(x)} (w_0(y) - \gamma_2 w_0(x)) \leq e^{w_1(y) - \lambda w_1(x)} - 1.$$

Proof. It is sufficient to show that there exists a constant $R_1 > 0$ satisfying

$$(53) \quad w_0(y) + \ln w_0(y) + K + \ln 2 + \eta w_1(x) \leq w_1(y) - \lambda w_1(x)$$

for all $y \in \mathcal{B}_{w_0}^c(\gamma_2 w_0(x))$, $x \in \mathcal{B}_{w_1}^c(R)$ and $R \geq R_1$. Note that for any $p \in (0, 1)$ and $\epsilon \in (0, 1)$, there exists a constant D (depending on p and ϵ) satisfying $x^p + p \ln x \leq \epsilon x + D$, $\forall x \geq 1$, which implies $w_0(x) + \ln w_0(x) \leq \epsilon w_1(x) + D$, $\forall x \in \mathcal{X}$. Hence, for all $y \in \mathcal{B}_{w_0}^c(\gamma_2 w_0(x))$, we have

$$\begin{aligned} & w_1(y) - w_0(y) - \ln w_0(y) - (\lambda + \eta)w_1(x) \\ & \geq (1 - \epsilon)w_1(y) - (\lambda + \eta)w_1(x) - D \geq \left((1 - \epsilon)\gamma_2^{1/p} - \lambda - \eta \right) w_1(x) - D. \end{aligned}$$

Choosing $\gamma_2 \in ((\lambda + \eta)^p, 1)$, $\epsilon < 1 - \frac{\lambda + \eta}{\gamma_2^{1/p}}$ and $R_1 := \frac{D + K + \ln 2}{(1 - \epsilon)\gamma_2^{1/p} - \lambda - \eta}$, (53) holds for all $y \in \mathcal{B}_{w_0}^c(\gamma_2 w_0(x))$, $x \in \mathcal{B}_{w_1}^c(R)$ and $R \geq R_1$. \square

LEMMA A.4. For any $\eta > 0$, $p \in (0, 1)$, $\gamma_2 \in (\lambda^p, 1)$ and $K \geq 0$, there exists a constant R_2 such that for all $y \in \mathcal{B}_{w_1}(\lambda w_1(x))$, $x \in \mathcal{B}_{w_1}^c(R)$ and $R \geq R_2$,

$$(54) \quad e^{-w_0(y) + \eta w_1(x) - K} (\gamma_2 w_0(x) - w_0(y)) \geq 1 - e^{w_1(y) - \lambda w_1(x)}.$$

Proof. It is sufficient to show that $e^{-w_0(y) + \eta w_1(x) - K} (\gamma_2 w_0(x) - w_0(y)) \geq 1$ under the same condition. Note that there exists a constant $D > 0$ such that

$$\frac{\gamma_2}{\eta} x^p \leq x + D, \forall x \geq 1,$$

which yields $-w_0(y) + \eta w_1(x) - K \geq -w_0(y) + \gamma_2 w_0(x) - K - D$ and hence,

$$e^{-w_0(y) + \eta w_1(x) - K} (\gamma_2 w_0(x) - w_0(y)) \geq e^{\gamma_2 w_0(x) - w_0(y) - K - D} (\gamma_2 w_0(x) - w_0(y)).$$

For all $y \in \mathcal{B}_{w_1}(\lambda w_1(x))$, we have $\gamma_2 w_0(x) - w_0(y) \geq (\gamma_2 - \lambda^p)w_0(x)$. Hence,

$$e^{\gamma_2 w_0(x) - w_0(y) - K - D} (\gamma_2 w_0(x) - w_0(y)) \geq e^{(\gamma_2 - \lambda^p)w_0(x) - K - D} (\gamma_2 - \lambda^p)w_0(x).$$

Due to the fact that $g(x) = e^x \cdot x$ is an increasing function on \mathbb{R}_+ , we can choose $\tilde{R}_2 > 0$ such that $e^{\tilde{R}_2} \cdot \tilde{R}_2 = e^{K + D}$. Hence, we have for all $y \in \mathcal{B}_{w_1}(\lambda w_1(x))$, $x \in \mathcal{B}_{w_0}^c(\tilde{R})$ and $\tilde{R} \geq \tilde{R}_2$, $e^{-w_0(y) + \eta w_1(x) - K} (\gamma_2 w_0(x) - w_0(y)) \geq 1$ holds. Finally, setting $R_2 = \tilde{R}_2^{1/p}$, the assertion is obtained. \square

Hence, by Lemma A.3 and A.4, for all $x \in \mathbf{B}_{w_1}^c(R_1 \vee R_2 \vee A)$ and $a \in \mathbf{A}(x)$,

$$\begin{aligned}
& \int_{\mathbf{B}_{w_0}^c(\gamma_2 w_0(x))} Q_{x,a}(\mathrm{d}y) e^{w_0(y)+K+\eta w_1(x)} (w_0(y) - \gamma_2 w_0(x)) \\
(\text{Lemma A.3}) \quad & \leq \int_{\mathbf{B}_{w_0}^c(\gamma_2 w_0(x))} Q_{x,a}(\mathrm{d}y) \left(e^{w_1(y)-\lambda w_1(x)} - 1 \right) \\
(52) \quad & \leq \int_{\mathbf{B}_{w_1}^c(\lambda w_1(x))} Q_{x,a}(\mathrm{d}y) \left(e^{w_1(y)-\lambda w_1(x)} - 1 \right) \\
(51) \quad & \leq \int_{\mathbf{B}_{w_1}(\lambda w_1(x))} Q_{x,a}(\mathrm{d}y) \left(1 - e^{w_1(y)-\lambda w_1(x)} \right) \\
(\text{Lemma A.4}) \quad & \leq \int_{\mathbf{B}_{w_1}(\lambda w_1(x))} Q_{x,a}(\mathrm{d}y) e^{-w_0(y)+\eta w_1(x)-K} (\gamma_2 w_0(x) - w_0(y)) \\
(52) \quad & \leq \int_{\mathbf{B}_{w_0}(\gamma_2 w_0(x))} Q_{x,a}(\mathrm{d}y) e^{-w_0(y)+\eta w_1(x)-K} (\gamma_2 w_0(x) - w_0(y)),
\end{aligned}$$

which implies that for all $f \in \mathcal{B}_{w_0}$ satisfying $|f| \leq w_0 + K$,

$$(55) \quad \int Q_{x,a}(\mathrm{d}y) e^{f(y)} (w_0(y) - \gamma_2 w_0(x)) \leq 0, \forall x \in \mathbf{B}_{w_1}^c(R_1 \vee R_2 \vee A).$$

Finally, for all $x \in \mathbf{B}_{w_1}(R_1 \vee R_2 \vee A)$ and $a \in \mathbf{A}(x)$ and $f \in \mathcal{B}_{w_0}$ satisfying $|f| \leq w_0 + K$,

$$\frac{Q_{x,a}[e^f w_0]}{Q_{x,a}[e^f]} \leq \frac{Q_{x,a}[e^{w_0+K} w_0]}{Q_{x,a}[e^{-w_0-K}]} \leq e^{2K} Q_{x,a}[e^{w_0} w_0] \cdot Q_{x,a}[e^{w_0}]$$

Using the fact that there exists some constant $D > 0$ satisfying

$$x^p + p \ln x \leq x + D, \forall x \geq 1,$$

we obtain that $Q_{x,a}[e^{w_0} w_0] \leq e^D Q_{x,a}(e^{w_1})$ which is upper bounded on $\mathbf{B}_{w_1}(R_1 \vee R_2 \vee A)$. Hence, there exists a $K_2 > 0$ such that for all $f \in \mathcal{B}_{w_0}$ satisfying $|f| \leq w_0 + K$,

$$\frac{Q_{x,a}[e^f w_0]}{Q_{x,a}[e^f]} \leq K_2, \forall x \in \mathbf{B}_{w_1}(R_1 \vee R_2 \vee A),$$

which together with (55) implies the required inequality.